Handbook on Cyclical Composite Indicators

FOR BUSINESS CYCLE ANALYSIS

IN COLLABORATION WITH
THE CONFERENCE BOARD, INC.

2017 edition

United Nations

MANUALS AND
GUIDELINES

eurostat

Handbook on Cyclical Composite Indicators

FOR BUSINESS CYCLE ANALYSIS | **2017 edition**

Foreword

The 2007-2009 global financial and economic crisis led to a further reflection on the need for an international agreed system of high frequency macroeconomic statistics and indicators. High frequency statistics and indicators based on an international agreed methodology would fill the data gap for determining an internationally comparable macroeconomic conditions and timely assessments of business cycle signals.

Such data and analytical gaps were extensively discussed in a series of international seminars (held in 2009 and 2010) jointly organised by the United Nations Statistics Division (UNSD) and Eurostat with the cooperation of Statistics Canada, Statistics Netherlands (CBS), and the Russian Federal State Statistics Service (Rosstat) with a broad range of statistical, analytical and policy stakeholders.

Based on the outcome of these international seminars, the United Nations Statistics Division and Eurostat, in close collaboration with Statistics Canada, Statistics Netherlands and Russian Federal State Statistics Service, developed an international programme on short-term economic statistics as part of a coordinated statistical response to the economic and financial crisis. This programme, which was endorsed by the United Nations Statistical Commission at its forty-second session in 2011, comprised of four thematic areas, namely: business cycle composite indicators, economic tendency surveys, rapid estimates, and data template for short term indicators and analytical indicators.

In the area of business cycle composite indicators, there was a consensus on the need for an internationally comparable set of cyclical composite indicators for business cycle analysis. At the same time, the active involvement of official statistical authorities was recognised, either in the direct compilation of such indicators or in the provision of critical support to compilers in other national agencies.

A Working Group led by CBS Netherlands was created and tasked with the preparation of the handbook on cyclical composite indicators, as a joint publication of Eurostat, The Conference Board, Inc. and the United Nations Statistics Division. This Working Group brought together a wide range of institutional expertise from international organisations, national statistical institutes, academic and research institutions as well as Central Banks.

More specifically, this handbook is intended provide statistical and econometric guidance on harmonized principles and methods for the compilation, monitoring and dissemination of cyclical composite indicators. It is expected that this handbook contributes to improved cross-country comparisons of economic performance using international agreed methodology based on international best practices.

This handbook is written for both producers and users of cyclical composite indicators. National statistical offices and other national producers that are considering a system of cyclical composite indicators will find sound methodological and practical guidance in setting up their production process. It provides recommendations on the choice of the reference cycle and on many other appropriate methodologies to be used.

Editors and Authors

Editors

Editor:	Mr. Gian Luigi Mazzi (formerly: Eurostat)
Co-Editor:	Mr. Ataman Ozyildirim, The Conference Board, Inc.
Technical Editor:	Mr. Dan A. Rieser, Eurostat

Working Group on cyclical composite indicators:

Statistics Netherlands:	Mr. Geert Bruinooge and Mr. Kees Zeelenberg–Chairs
Eurostat:	Mr. Gian Luigi Mazzi
The Conference Board, Inc.:	Mr. Ataman Ozyildirim
United Nations Statistics Division:	Mr. Ivo Havinga, Mr. Herman Smith and Ms. Ilaria DiMatteo

Contributing Authors (in alphabetical order)

Nadim Ahmad
OECD
nadim.ahmad@oecd.org

Jacques Anas
Independent Expert
jacques.anas@free.fr

Roberto Astolfi
OECD
roberto.astolfi@oecd.org

Monica Billio
Ca' Foscari University of Venice
billio@unive.it

Ludovic Calès
European Commission, Joint Research Centre (JRC)
ludovic.cales@ec.europa.eu

Gonzalo Camba-Mendez
ECB
gonzalo.camba-mendez@ecb.europa.eu

Leonardo Carati
GRETA
leonardo.carati@gmail.com

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

Editors and Authors

Françoise Charpin
L'Institut d'Études Politiques de Paris (Sciences Po)
françoise.charpin@sciences-po.fr

Laurent Ferrara
Banque de France
laurent.ferrara@banque-france.fr

Michela Gamba
OECD
michela.gamba@oecd.org

Dominique Guégan
Université Paris 1 Panthéon-Sorbonne
dominique.guegan@univ-paris1.fr

Emmanuelle Guidetti
OECD
emmanuelle.guidetti@oecd.org

Gyorgy Gyomai
OECD
gyorgy.gyomai@oecd.org

George Kapetanios
King's College, University of London
george.kapetanios@kcl.ac.uk

Sabine Krieg
Statistics Netherlands (CBS)
s.krieg@cbs.nl

Matteo Luciani
Board of Governors of the Federal Reserve System
matteo.luciani@frb.gov

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Massimiliano Marcellino
Bocconi University, IGIER, and CEPR
massimiliano.marcellino@unibocconi.it

James Mitchell
Warwick Business School, University of Warwick
james.mitchell@wbs.ac.uk

Filippo Moauro
ISTAT
moauro@istat.it

Pim Ouwehand
Statistics Netherlands (CBS)
p.ouwehand@cbs.nl

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

Fotis Papailias
King's College, University of London and quantf research
f.papailias@kcl.ac.uk

Claudio Parracho-Soares
European Commission, Eurostat
claudio.parracho-soares@ec.europa.eu

Pierre-Alain Pionnier
OECD
pierre-alain.pionnier@oecd.org

Dan A. Rieser
European Commission, Eurostat
dan.rieser@ec.europa.eu

Rosa Ruggeri Cannata
European Commission, Eurostat
rosa.ruggeri-cannata@ec.europa.eu

Marc Smeets
Statistics Netherlands (CBS)
mset@cbs.nl

Frank van de Pol
Statistics Netherlands (CBS)
f.van.de.pol@home.nl

Jan van den Brakel
Statistics Netherlands (CBS)/Maastricht University
j.vandenbrakel@mastrichtuniversity.nl

Floris van Ruth
Statistics Netherlands (CBS)
frth@cbs.nl

Piet Verbiest
Statistics Netherlands (CBS)
pvrt@cbs.nl

Martin R. Weale
King's College, University of London
martin.weale@kcl.ac.uk

Table of Contents

1

Introduction, Aim and Scope

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

Table of Contents

1.1 Introduction

The 2007-2009 global financial and economic crisis, the so-called Great Recession, revealed, among other things, some weaknesses of the system of macroeconomic infra-annual statistics at global level. These weaknesses have constituted one of the factors preventing a prompt and effective detection of the crisis signals. Furthermore, the infra-annual system of macro-economic statistics was unable to promptly deliver the expected feedback and warnings helping policy makers to quickly react to mitigate most of the crisis effects. Obviously, the situation of infra-annual statistics was quite different across countries and regions but some weaknesses were, in a more or less pronounced way, common to all of them.

The main observed weaknesses could be synthesised as follows:

– the lack of an homogeneous set of statistical indicators being able to provide a reliable overall picture of the macroeconomic situation and also allowing for cross-country comparisons;

– the lack of timeliness observed for most of key macroeconomic indicators and in particular for the National Income and Product Accounts, and particularly GDP; and

– the difficulties encountered by users and analysts in extracting cyclical signals from official statistics and the consequent lack of new (maybe composite) indicators providing a clear picture of cyclical movements and of the occurrence of turning points.

Starting from the above considerations, the United Nations Statistical Division (UNSD), the statistical office of the European Commission (Eurostat) and the Centre Bureau of statistics of the Netherlands (CBS) launched a joint initiative in response to the global financial and economic crisis. This initiative led to the organisation of a series of three international seminars, held in 2009 in Canada and the Netherlands and in 2010 in the Russian Federation. A wide range of statistical agencies from all around the world, as well as central banks, research institutes and academics were taking part to these meetings, providing their own experience and actively contributing in the formulation of proposals.

The topic of the first international seminar was "Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends" while the second and the third ones dealt with "Early Warning and Business Cycle Indicators". During the three seminars, the state of infra-annual macroeconomic statistics was subject to an in-depth analysis and investigation targeting the identification of the main weaknesses of the hole system of macro-economic infra-annual statistics. After a long and productive discussion, accompanied by the presentation of several interesting papers from different statistical agencies, as well as central banks and key users, some corrective measures were proposed.

The main corrective measures suggested as the outcome of this joint initiative were the following:

– the preparation of a glossary of rapid estimates in order to achieve a consensus around a common terminology associated to various kinds of rapid estimates produced by statistical agencies;

– the preparation of a handbook of rapid estimates providing a methodological overview of statistical and econometric techniques useful to foster timeliness of cyclical statistics as well as some guidance on the construction of rapid estimates;

– the preparation of a handbook on cyclical composite indicators (CCI) presenting the various kinds of composite indicators and the associated statistical and econometric compilation techniques and methods as well as some guidance for their construction;

– the preparation of a handbook on opinion tendency surveys complementing the already existing ones and providing more advanced compilation guidance for this essential piece of information for short term economic analysis; the preparation of a handbook on data template and metadata for short term economic statistics providing guidance for the development of a common set of macroeconomic indicators for monitoring the economic situation with associated standardised metadata.

The overall objective of those initiatives was the enhancement of the set of available data and tools aiming to provide an effective monitoring of the short-term economic situation. Such an effective monitoring could eventually be implemented as a real-time economic and cyclical early warning system. The idea of an effective monitoring of the short-term macro-economic situation was presented in Mazzi and Montana (2009) and an attempt to design a cyclical early warning system was proposed by Mazzi et al. (2011).

The present handbook — a publication prepared jointly by Eurostat, The Conference Board and the United Nations Statistics Division — is part of the set of the handbooks listed above. Finally, it is worth noticing that this handbook does not aim to influence statistical agencies on their decision to compile or not composite indicators. It just aims to provide methodological support and guidance either to statistical agencies already compiling cyclical composite indicators or to those envisaging their compilation in the near future, or even to the ones thinking to play an active role in the compilation of cyclical composite indicators by other agencies. This introductory chapter is structured as follows: Section 1.2 contains some general considerations about composite indicators while Section 1.3 presents the main aim, scope and limitations of this handbook; section 1.4 shortly details the structure of the handbook and section 1.5 shortly concludes.

1.2 General considerations on composite indicators

The expression "composite indicators" is widely used in the literature and in empirical applications, and sometimes it is even abused. There is no unique meaning for this expression. This is the main reason for which here we are trying to better clarify the concept of composite indicators and within it the one of cyclical composite indicators.

There is no doubt that, strictly speaking, some, perhaps most, key statistical indicators can be viewed as composite indicators since they stem from an aggregation of component series based on a given weighting scheme. This is the case, for example, of the GDP, inflation, employment and unemployment, etc. Nevertheless, those indicators are derived by aggregating information coming from their own statistical domain (e.g. national accounts, price statistics, labour market statistics) and using a fixed weighting scheme based on clear statistical principles (weights usually reflect the relative importance of each component series). By contrast, composite indicators and their component series often come from different statistical domains and are not necessarily measured in the same units. While the former set of statistics relate directly and closely to singular economic concepts such as employment or economic output, the latter, composite indicators, are related to concepts that are not directly observable or higher level concepts such as economic performance or economic activity. Furthermore, the weighting scheme, if any, reflects more subjective considerations or some statistical properties of the component series in terms of forecasting ability, etc.

Finally, composite indicators aim to extract, in real time or for the near future, relevant signals, sometimes unobserved (e.g. cyclical ones) or latent (not directly measurable phenomena) such as poverty, etc. Composite indicators play also a quite different role in socio-economic areas and in the macro-economic and financial ones. In the former they are often criticised, mainly for the fact of applying statistical tools and techniques to often qualitative unmeasurable phenomena. An analysis of the main difference of composite indicators in socio-economic and macroeconomic areas is presented in Mazzi et al (2016). In macroeconomics and finance, composite indicators are a very relevant complement to the traditional statistical information system since they can provide very useful signals of risks, such as cyclical recessions or slowdowns, instability due to macroeconomic imbalances of sectoral stress, etc. Composite indicators, in the macroeconomic and financial fields, can be either related to cyclical aspects or to more structural ones. This handbook is focusing on the former category for which a detailed taxonomy is proposed by Carriero et al. (2017) (Chapter 3).

1.3 Aim, scope and limitations of the handbook

This handbook aims to provide an overview of various types of cyclical composite indicators as well as of the most commonly used compilation techniques and methods. It also provides a methodological overview of the compilation approaches and it describes, as far as possible, their advantages and drawbacks. Various topics are presented in a didactic manner, making the reading easier for both non-expert and skilled users. Complex formalisations are also included in a way that they do not affect the understanding of the techniques by beginners and non-expert users. The scope of the handbook is limited to the cyclical composite indicators so that other kinds of composite indicators also used in macro-economics are not discussed here.

Furthermore, the handbook, describes (also providing detailed selected readings and references) mostly well assessed and consolidated methodologies so that the most recent developments, still in an experimental phase or restricted to the academic world, are not considered yet. Finally, structural models which imply the adoption of one or another economic theory are also not considered. We focus our attention mainly on data driven methods or on methods based on statistical evidence. Nevertheless, it is clear that compilers of cyclical composite indicators should pay special attention on the fact that their indicators should be, not only statistically, but also economically sound, avoiding the use of any implausible relationship. The target readers of the handbook are mainly official statisticians and central bankers but also researchers, students and academics can profit from the content of this handbook.

1.4 Structure of the handbook

The handbook is structured in seven parts, each of them composed of two or more chapters drafted by internationally acclaimed authors and experts in the field.

- Part I contains an historical overview of the business cycle theory accompanied by an analysis of the various ways of constructing cyclical indicators since the 19th century. Part I also contains a detailed taxonomy and classification of cyclical composite indicators accompanied by some first empirical considerations on their compilation.

- Part II deals with all statistical and modelling aspects which need to be considered prior to the compilation steps. In particular, chapter 4 (van den Brakel et al. (2017)) introduces all problems related to the utilisation of official statistics in constructing cyclical composite indicators focusing on aspects such as the back calculation methods to obtain long and consistent time series, the construction, the use and the comparison of index series, the seasonal and calendar adjustment, etc. Chapters 5 and 6 (Guégan (2017) and Kapetanios G., Marcellino M. and Papailias F. (2017), respectively) deal with variable and model selection methods providing a general overview of the most widely disseminated methods and tools which can also be used when building up cyclical composite indicators.

- Part III describes the most widely used compilation techniques for coincident and leading indicators aiming to estimate in real-time or anticipate cyclical movements.

- Part IV describes the so-called turning points composite indicators, also known as recession indicators. This part also introduces in some details the various typologies of non-linear regression and time series models which could be used in this context.

- Part V deals with those composite indicators aiming to estimate in real-time or anticipate the short term behaviour of a given economic indicator such as GDP, unemployment, etc. These indicators constitute an alternative to the rapid estimates techniques to improve the timeliness of macro-economic statistics and, for this reason, this part can be also of interest for readers mainly interested in in issues related to the enhancement of the data production timeliness. Part III, IV and Part V start with an introductory chapter providing overview of the most relevant topics discussed in the session guiding the reader

through the contents various chapters. Its introductory chapter provides some useful bibliographical references as well las some information about already existing composite indicators.

– Part VI discuss some practical aspects relate to the performance of cyclical composite indicators and on possible attractive ways to disseminate their outcomes.

– Finally, Part VII, composed of a single chapter, contains some compilation guidelines and best practices recommended when starting the compilation of new cyclical composite indicators or for reviewing existing ones.

1.5 Conclusions

In this short introductory chapter, we have presented the historical background which led to the decision of preparing this handbook and we have also presented the philosophy we followed in its preparation, which has obviously influenced its aim, scope and limitations. There is no doubt that cyclical composite indicators can constitute a very useful complement to the information provided by traditional sources of information such as official statistics. They can contribute to highlight hidden components but also to provide timelier estimates of the key macroeconomic indicators.

In this respect, cyclical composite indicators appear to be an essential component of a real-time economic monitoring and cyclical early warning system, which could ensure a more reliable and effective responses to the current economic situation by policy makers and decision makers. Even if it is not realistic to expect preventing all future new crises, these new tools could contribute to a quicker reaction of policy makers in order to mitigate effects of future crises and maybe to shorten their consequences in time and costs. In a globalised world, the availability of comparable cyclical composite indicators across countries and regions could be a further element in strengthening the implementation of quick and effective counter-cyclical measures, coordinates across countries. This handbook could contribute to the achievement of these goals in the next years. Obviously the contribution provide by this handbook is limited to technical and methodological aspects as well as to the definition of best practices. All aspects related to the political aspects of the compilation of cyclical composite indicators remain out of the scope of this handbook.

Bibliography

van den Brakel J., Krieg S., Ouwehand P., van de Pol F., van Ruth F. and Smeet M. (2017). Chapter 4: Composite Estimators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Carriero A, Marcellino M. and Mazzi G.L. (2017). Chapter 3: Definitions and Taxonomy of indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Guégan D. (2017). Chapter 5: Review of Parametric and non-Parametric Variable and Model Selection Techniques, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Kapetanios G., Marcellino M. and Papailias F. (2017). Chapter 6: Variable Reduction and Variable Selection Methods using Small, Medium and Large Datasets: A Forecast Comparison for the PEEIs, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mazzi G.L., Moauro F. and Ruggeri-Cannata R. (2011). A statistical overview of the economic situation in the euro area, 46th Scientific Meeting of the Italian Statistical Society.

Mazzi G.L and Montana G. (2009). A system of rapid estimates to improve real-time monitoring of the economic situation: the case of the Euro Area, National Institute Economic Review n.210, October 2009, 63-70.

Mazzi G.L., Mitchell J., and Moauro F. (2016), Structural vector autoregressive (SVAR) based estimates of the euro area output gap: theoretical considerations and empirical evidence, Eurostat Statistical Working Papers, Publications Office of the European Union.

Historical Overview and Definitions

2
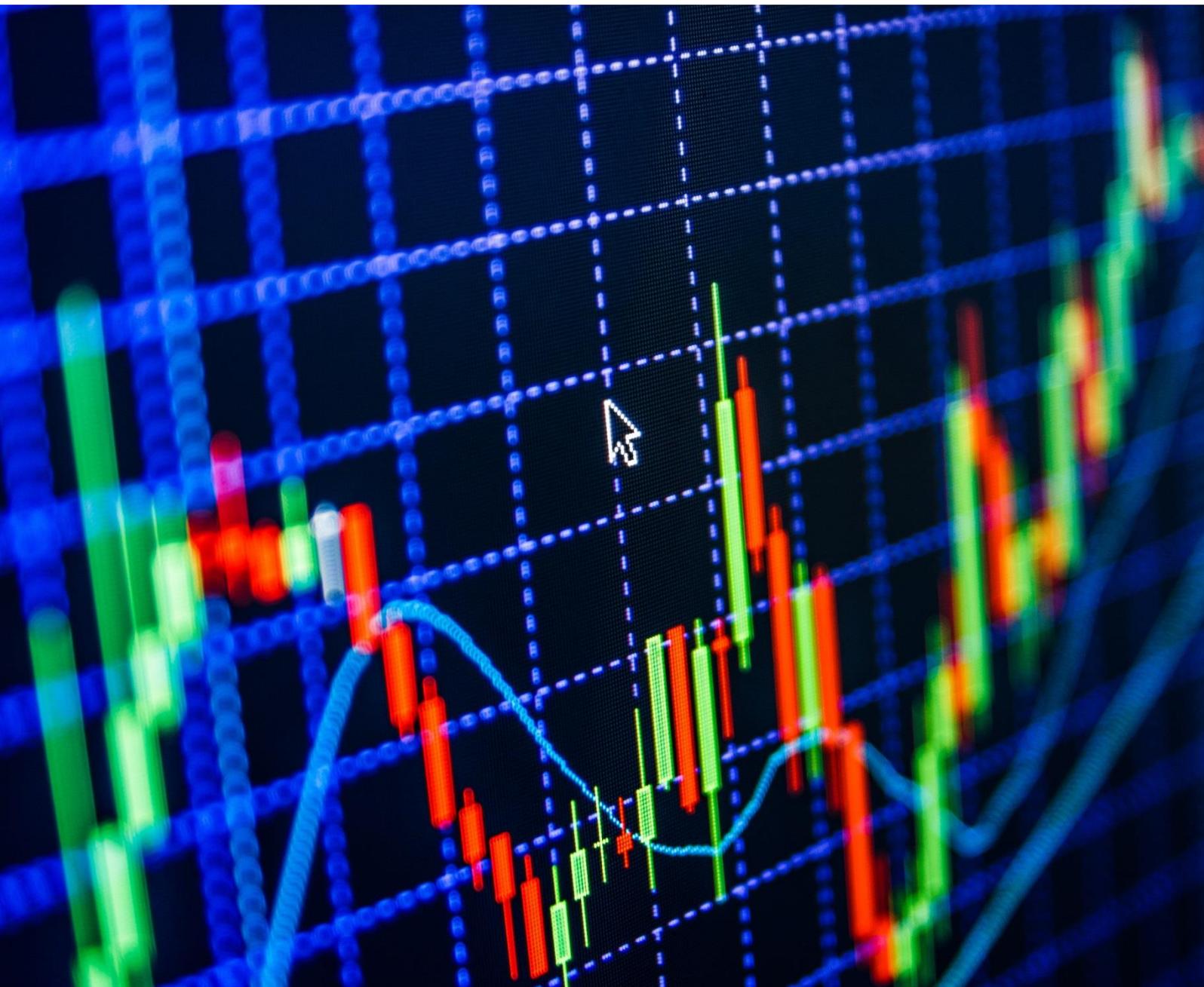
**Business Cycles Theories:
An Historical Overview**
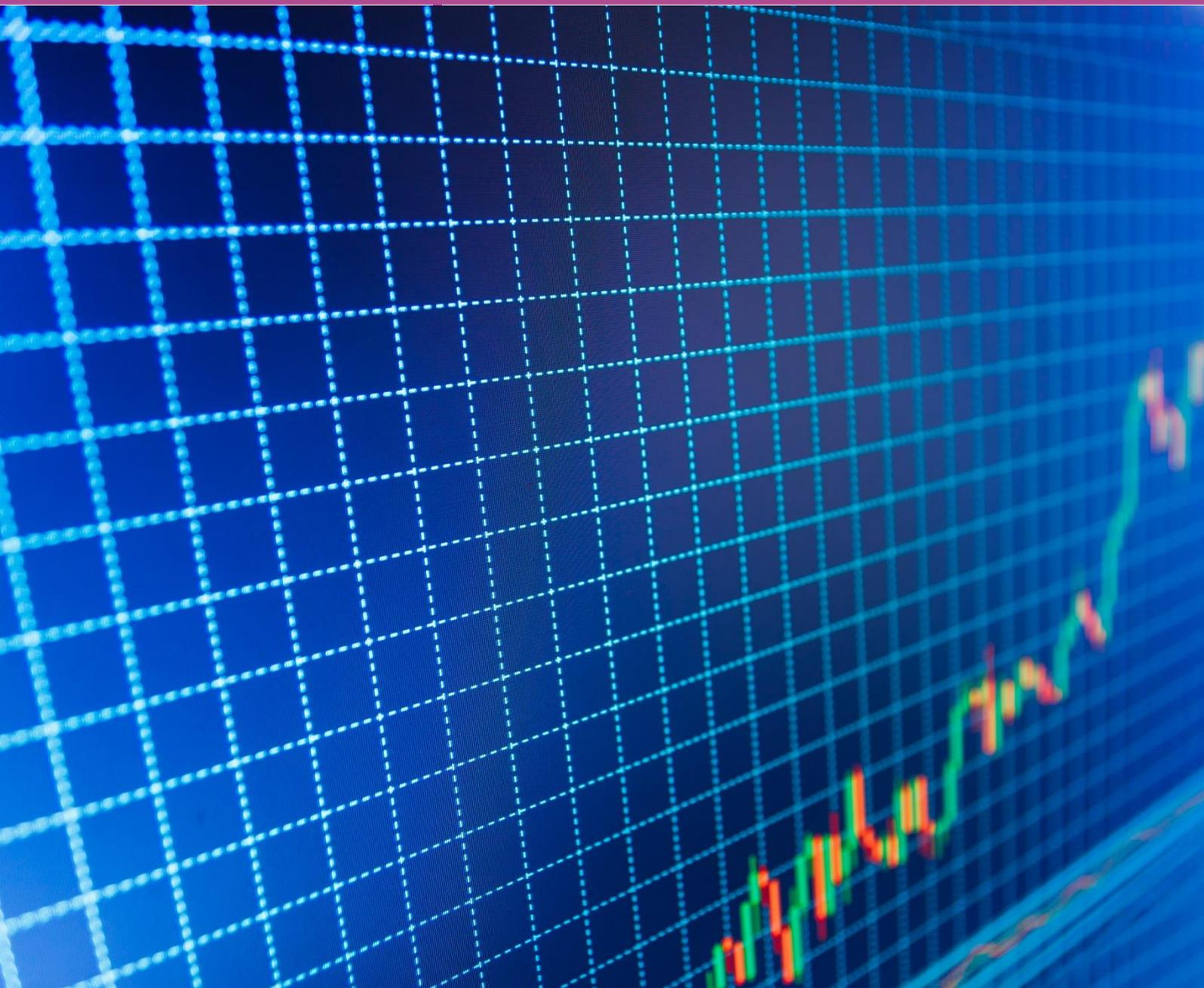
Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

Table of Contents

2.1 Introduction

The aim of this chapter is to provide an historical overview of the evolution of the theories of economic cycles since the end of 19th century as well as of the tools developed for cyclical monitoring diagnosis and prognosis. In this chapter we are trying to alternate, following as much as possible an historical sequence, the presentation of theories and tools to give an idea of how both were evolving and in what direction in a given period of time. The idea that the aggregate economy does not climb with steady trend but experiences occasional booms of activity and recessions is very old. Virtually every economist recognises the existence of fluctuations in the general level of economic activity. But the idea that the whole economy exhibits a regular cyclical pattern, meaning that these fluctuations were recurrent in a precise periodic way, was only put forward late in the 19th century. In this chapter we are not following any precise classification of theories of economic cycles but we are presenting them in temporal sequence having the Keynesian revolution as the reference point. Several interesting classifications have by the way been proposed in the literature. Here we would like to mention three of them which could be beneficial to readers of the following sections. The first one makes the difference between:

- Fully exogenous or natural theories where cycles are determined by some natural phenomenon;

- Fully endogenous theories, where cycles are generated by causes internal to the economic system only;

- Theories based on some shocks propagation mechanism where shocks can occur also outside the economic system (exogenous shocks).

Another, possibly even more interesting, classification distinguishes between:

1. self-correcting theories (mainly the classical theories);

2. non self-correcting theories allowing for active countercyclical policy interventions (Keynesian approaches);

3. policy ineffectiveness based theories (the neo-classical ones);

4. contract-based wage and price stickiness theories (the new-Keynesian ones).

Concerning the tools development, we can observe that this was quite moderate until the end of 40s with not very strong relations with cyclical theories. By contrast, after, there was a remarkable development of tools, mainly based on econometric methods and on time-series analysis. Tools also started to be much more related to cyclical theories.

The structure of the chapter is as follows: Section 2.2 describes fully exogenous or natural theories of business cycles. Section 2.3 outlines different, existing economic barometers, whilst section 2.4 describes different pre-Keynesian theories of cyclical fluctuations. Section 2.5 outlines the Cambridge school of thought's view on the matter, section 2.6 the Keynesian theory. The empirical revolution brought about by Burns and Mitchell is presented in section 2.7. Section 2.8 presents other definitions of cycles, such as the growth cycle, the acceleration cycle and the unified framework. Section 2.9 presents contemporaneous theories of business cycles, whilst section 2.10 describes the evolution of business cycle tools. Section 2.11 synthesises new challenges for business cycle analysis arising after the global economic and financial crisis. Finally, section 2.12 concludes.

2.2 Fully exogenous or natural theories

The first theorists of economic fluctuations provided mostly "fully exogenous' justification cycles". These theories were relating economic cycles to other exogenous cyclical phenomena found in "nature" such as weather, which in turn might be affected by astral phenomena. These theorists argued that natural phenomena

affected tangible things such as harvests and/or intangibles such as people's moods, and that these in turn were generating the observed economic fluctuations. As these natural phenomena had a cyclical nature, so then we see a corresponding economic cycle. However, the industrialising nations of the 19th century seemed to exhibit a different type of regular cyclical behaviour that could not be directly attributed to natural factors. These fluctuations became known as trade or business cycles.

In this section we are mainly focusing in two theories proposed between the end of the 19th and the 20th century.

2.2.1 Jevons' sunspots theory

William Stanley Jevons (Jevons (1866),Jevons (1875),Jevons (1884)) argued that sunspots, or intense solar activity that occurs periodically (approximately every 10 to 12 years), cause variations in the climate, alternatively generating periods of drought and flooding.

These periods in turn cause booms and busts in the agriculture industry, which then cause fluctuations in overall economic activity. In modern industrialised economies a sunspot driven economic cycle would have little impact on the overall level of economic activity. In lesser developed and in the agrarian economies the causal relationship could possibly have more impact. His naive explanation could not long withstand critical examination. It had a certain interest, however, in suggesting a causal factor that was completely detached from the economic system and one that could not be influenced by it in turn.

2.2.2 Moore's Venus theory

Later on, Henry L. Moore proposed a theory based on weather cycles (Moore (1914)) and another on the position of the planet Venus (Moore (1923)). At the beginning of the thirties, Akerman (1928), Akerman (1932) had a more ingenious one connecting longer business cycles to the magnified effects of a series of small, weather-driven seasonal cycles.

2.3 The economic barometers

Until the first war, we can consider that a real assessment of the economic fluctuations was not properly defined. The literature was essentially based on exotic studies without any theoretical foundation.

A significant improvement has then been made with the development of the so-called "economic barometers", which can be considered the first attempt to analyse and monitor economic fluctuations.

The "economic barometers" are directly derived from the instrument used in meteorology to measure atmospheric pressure to assess the evolution of the weather. Economic barometers constitute the first attempt of providing a regular assessment of the economic situation and of anticipating future macro-economic movements hopefully to avoid further economic crises. While the largest diffusion of economic barometers characterises the periods between the two world wars, first examples of barometers already appeared at the end of the 19th century. Probably, as Armatte (2003) stated, the main reasons for the development of economic barometers at that time is due to two main reasons: the first one is related to the fact that at the end of 1880's crises started to be more recurrent and generalised than before; the second is related to the fact that at this time, economic statistics started a remarkable development. A very complete review of economic barometers is provided byArmatte (1992), Armatte (2003) and in this chapter we are relying on those contributions to provide a very synthetic overview of economic barometers. We can consider 3 different classes of economic barometers based on their construction principle, even if this classification does not necessarily reflect the chronology of their construction and publication:

- Simple barometers

- Multiple or composite barometers

- Synthetic barometers

Simple barometers

These barometers were constructed under the hypothesis that just one single indicator can appropriately describe economic fluctuations. Such selected indicator was presented in a graphical or tabular form after very elementary transformations. This type of barometers has been gradually abandoned after 1914. The main reason for abandoning them was that just one single variable cannot synthesise all economic movements. Consequently, the risk of delivering misleading signals was considered too high by analysts. Furthermore, the economic system is continuously evolving over the time and structural changes were considerably modifying the role and the importance of various economic activities and related indicators.

As a matter of fact, the nature of economic crises was essentially agricultural in the XVIII and at the beginning of XIX century. Then they become essentially commercial, industrial and, more recently, energetic and financial. The most known simple barometers are:

- William Farr barometer using the number of weddings as an indicator of the economic activity (Farr, William (1769)).

- Juglar barometer using the portfolio situation of the French central bank or its liquidity (which was moving in the countercyclical way with respect to the first one) as an indicator of the economic movement (Juglar (1862)).

- Mourre (1913) barometer using the rate of commercial paper in New York as the best indicator of the monetary situation.

Multiple or composite barometers

This family of barometers is based on a quite high number of indicators, which are not yet combined or synthesised. Such barometers were essentially used to assess the current situation of the economy and their utilisation to anticipate future movements was quite limited. The two predecessors of this family of tools were the Neumann-Spallart (1887) and De Foville (1888) barometers. Those two very similar barometers were intended not only to assess the economic situation of a given country but also to allow for an international comparison of the economic conditions. They were based on a large number of indexes classified into three groups representing: the state of the economy, the social condition, the moral state of society. After those pioneering barometers, probably the most known example of multiple barometers is proposed by Beveridge (1909). This barometer is based on a table called "The pulse of the nation" containing 13 indicators which cover a quite wide range of social and economic activity for the UK during the period 1856 to 1907.

The Belgian barometer published by Julin (1913) can be considered as a bridge from this family of barometers to the synthetic barometer.

While multiple barometers recognise that the assessment of the economic situation cannot be done by using just a single indicator, they were characterised by some drawbacks which limited considerably their practical utilisation. The first drawback was represented by the very large number of indicators used which was somewhat preventing the possibility of detecting clear and univocal detection of signals. The second one was represented by the heterogeneity of the indicators included in the barometers covering quite disparate areas not necessarily strongly related to the economy. The last one is related to the difficulty to use those barometers for forecasting purposes mainly due to the inadequate level of development of statistical tools available at that time, especially concerning multivariate analysis.

Synthetic barometers

The aim of synthetic barometers is to combine and synthesise a homogenous set of indicators representing a specific economic sector or a particular aspect of the economic fluctuations. The aggregation/synthesis is

obtained by means of statistical criteria. For the first time, series are chosen on the basis of some objective criteria such as their degree of similarity. The outcome can be either a synthetic index or a small number of synthetic indices whereby each of them reflects the behaviour of a specific economic sector or cyclical movements based on the selected subset of the chosen basic indicators. The most known synthetic barometers are:

- A) Babsoncharts barometer
- B) Harvard barometer
- C) European barometers.

A) Babsonchart barometer

This barometer was mainly developed for the need of private enterprises by a consultancy company led by Roger W. Babson. This barometer was disseminated by means of a weekly publication to the subscribers only. The barometer was based on 12 statistical series, of which 4 came from the monetary sector, 4 from financial sector and 4 from the commercial sector. Series were then aggregated by simple addition after a transformation in a common base. The fluctuations of this index were presented in a graphical way where an arbitrary drawn line was intersecting them. This line was drawn in such a way that the surfaces above and below it were almost equivalent (see figure 2.1, from Armatte (2003)). The main difference with respect to most of the already presented barometers is that the Babsonchart is mainly constructed with a forward-looking purpose, even if the forecasting tools used were very basic.

Figure 2.1: A Babson Chart
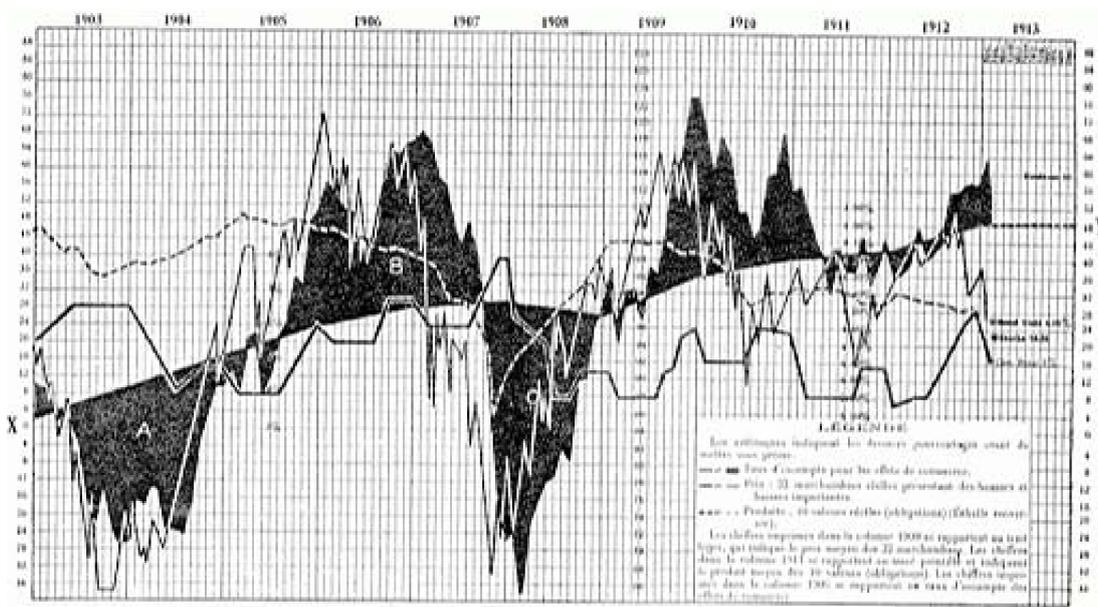
B) Harvard barometer

This barometer is likely the most advanced and modern among those developed during the period before the so called "Great Depression of 1929". It is characterised by an intensive use of new statistical tools which were giving it a higher degree of robustness with respect to the previous ones. The Harvard barometer was developed by Persons (1916), Persons (1919). The barometer was constructed originally from 20 indicators selected out of 50 available on government publications. Such indicators were then trend and seasonally adjusted. The resulting cyclical components of the indicators were then normalised to the same dispersion unit by dividing them by their respective standard deviations. The series so obtained were then analysed in detail in order to identify their relationships, including their leading/lagging structure. The first investigation was mainly based on a graphical analysis and on subjective appreciation, where in a second phase the

contemporaneous and shifted correlations were computed in order to have a clearer idea of the structure of the data. Finally, the information was synthesised into three curves representing the movements respectively in financial markets, goods markets and money market. In the Harvard[b] barometer, the movements of the first curve are anticipating those of the second one which anticipate those of the money market, so that the financial market was considered the leading indicator able to anticipate economic crises (see figure 2.2, from Armatte (2003)).

Figure 2.2: The Harvard barometer
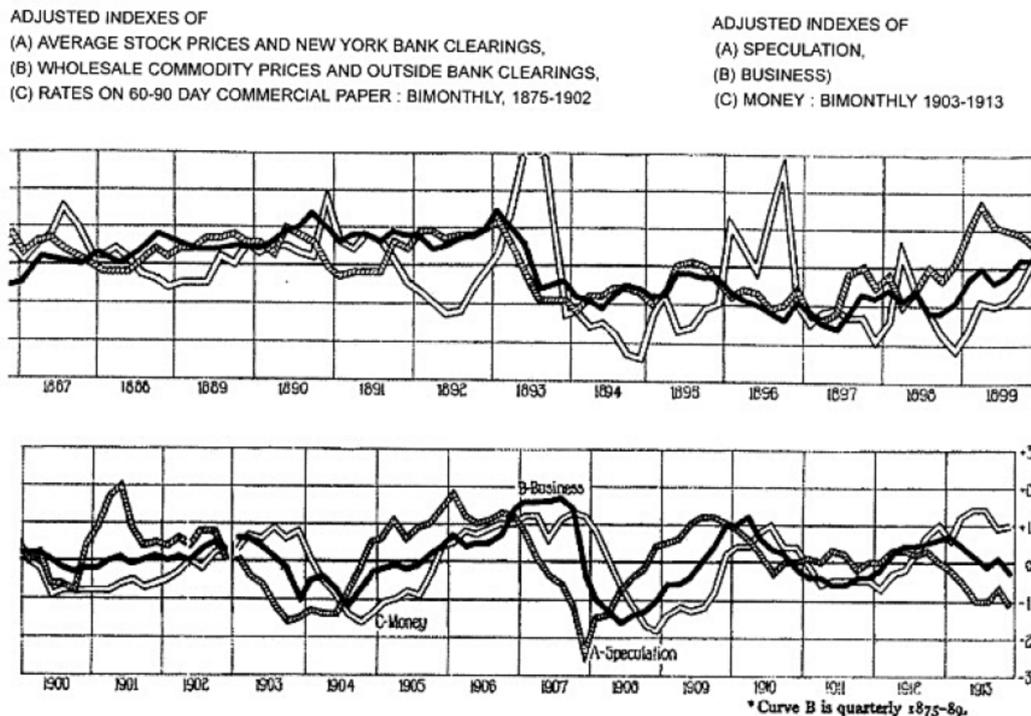
The Harvard barometer performed quite well in anticipating the 1920 crisis but it failed dramatically in anticipating the big crisis of 1929 and this event can be considered the end of the barometers era. Despite the unsuccessful results in 1929, the Harvard barometer can be, for several reasons, considered as the real precursor of modern composite indicators. The first reason is represented by the use of the correlation to identify the leading/lagging structure of the variable. The second reason is that for the first time, in the Harvard barometer, the seasonality was systematically removed from the component series before the analysis by using the seasonal adjustment method developed by Persons himself. Finally, in the Harvard barometer, series were regularly detrended by using linear or polynomial trend specifications. Those elements show some similarities with more modern and advanced cyclical composite indicators especially those having the growth cycle as the reference.

C) European barometers

Starting from the success of the Harvard barometer, several similar products were developed in Europe, for example by Beveridge in UK, March in France, Dupriez in Belgium, Wagemann in Germany and Kondratiev in Soviet Union.

2.4 Pre-Keynesian theories of cyclical fluctuations

Beside the large variety of studies aiming at identifying, measuring and, whenever possible, forecasting cyclical movements (see section 2.3), the end of the 19th and the beginning of the 20th centuries were also characterised by a huge literature about the causes determining the business cycle. This literature mainly aims

at identifying causes of cyclical fluctuations and to explain them within a given macro-economic theory. This section is an attempt to briefly summarise the most relevant pre-Keynesian theories of cyclical fluctuations. This review is far from being exhaustive and intends just to provide a general view of economic thoughts in relation to cyclical fluctuations.

The first attempts to explain cyclical movements did not have strong economic foundations, mainly because they were based on exogenous theories (see section 2.2). At the same time, with a few exceptions, the authors of barometers (see Section 2.3) did not investigate in detail the nature and the causes of the cycles they were trying to monitor and anticipate. On the other hand, the second part of the 19th and the first part of the 20th century were characterised by a proliferation of studies aiming at interpreting cyclical fluctuations in the light of economic theories and to identify causes of cyclical fluctuations. A quite comprehensive review of these studies can be found in Fayolle (2003).

In this section, we are providing a review of the most relevant business cycle theories before the Keynesian revolution. It is worth mentioning that in this historical period, for the first time, the dichotomy between deterministic, or fully endogenous, theories of cyclical fluctuations, and stochastic theories, with some degree of exogeneity emerged. This distinction characterises the economic debate even today.

2.4.1 Under-consumption Economic Cycles

Those two authors can be considered to be the first having developed this theory which was further receiving a great attention by several economists. They argued that the cause of the upper turning point is the decreasing ability of the economy to continue consuming what it produces during the expansion phase.

More than a theory, the under-consumption is a framework in which several theories based on this concept have been developed. All those theories were essentially developed in opposition to the classical economic view. An interesting overview of under-consumption theories is presented by Allgoewer (2002). She mainly focuses on the under-consumption theories developed at the beginning of the 20th century by John Hobson (1933), Foster and Catchings (1925).

Marx's view of instability has strong under-consumption connotations. His analysis led him to predict that the revolution would come first in the most mature capitalist country. The aggregate demand theory developed by Keynes is also based on some aspects of under-consumption theories. Nevertheless, in the Keynesian view, an under-consumption situation does not necessarily lead to an economic crisis since, within the aggregate demand theory, under-consumption can be counterbalanced by other components such as growth, fixed capital formation and government consumption.

2.4.2 The savings-investment theories

These theories contemplate the impact that the existence of an elastic monetary system could have on the economic equilibrium. These theories differ from classical ones mainly in the way in which the relationships between the banking system and the investment are designed.

The market law and the assumption of flexible wages and prices led to the conclusion that full employment is the long-term natural equilibrium for the economy. Business fluctuations were minor self-correcting phenomena. The cyclical expansion of the money supply is caused by profit-seeking bankers. Therefore, Wicksell (1898, 1936) introduced the distinction between natural and market interest rates.

Under the natural interest rate (in), the supply of savings would equal the demand for loanable funds, which was determined by what entrepreneurs wished to invest. Market interest rate (im), equated the demand for funds with the bank-inflated supply of loans. Hence, the impact of the operation of the banking system was to drive the market interest rate below the natural rate. Only when the two rates are the same the economy achieves the equilibrium. The divergence between the two interest rates generates cyclical fluctuations due

to the fact that the investment demand is not balanced by an adequate saving availability. A more detailed investigation of those theories can be found in Cottrell (1993) and Festré (2009).

Monetary over-investment theory: Friedrich von Hayek

In particular, von Hayek raises the question: Why is it that producer goods industries exhibit much greater amplitude over the cycle than consumer goods industries? He argued that the banking system makes loans available to entrepreneurs during the upswing, causing the market rate of interest to fall below the natural rate.

Hayek's major contribution was to claim that the banks lent to the higher stages of production, thereby enabling them to expand. If bank credit enabled the higher stages to bid resources away from the lower stages, expansion would lead to higher income, whereas consumer goods output would not have increased - and might even decrease. This situation would necessarily lead to an increase in consumer prices. The result would be higher prices for consumer goods, and with higher prices, consumers could not buy as much out of any level of nominal income. This is called forced saving.

The theory is based on a simplistic view of how banks perform. Its weaknesses would include its failure to appreciate that real, as well as monetary factors, can lead to capital formation (changes in profit expectations, technological changes, etc.). In general, it has been criticised for over-emphasising the importance of monetary factors (changes in interest rates) and its relatively unrealistic view of the investment process. Hayek pays no attention to the impact of changes in expectations on the behaviour of entrepreneurs. Hayek's theory, while offering an interesting rationale for the fixed investment cycle, can be considered an inadequate explanation of modern business cycles because it ignores too many critical factors.

2.5 Cyclical theories stemming from the Cambridge School of thought

2.5.1 Purely monetary theory of Hawtrey

Purely monetary theory proposed by Hawtrey (1913), Hawtrey (1919), and Hawtrey (1928) is based on quantity theory of money derived from the Marshallian or Cambridge equation of exchange. This equation is $M = k * Y$ where M is cash balances, k is a positive constant, Y is nominal income. This equation is the basis for the modern approach to monetary instability.

In its most rigid version, in which prices are assumed sticky, the money supply is exogenous, and k is fixed, it provides a precise guide to monetary policy. It also provides support for activities of central banks. Changing the conditions in the supply of money is assumed after some period to move nominal economic activity in the same direction but not by a precisely specifiable amount. This theory exaggerates the importance of the traders, mainly their sensitivity to interest rate changes. It failed to realise that business fluctuations are likely to be more than monetary phenomena.

The Hawtrey thought is well synthesised by the following quotation from his book Good and Bad Trade (1913):

"The real starting-point of the whole [argument] is to be found in the thesis. . . that a depression of trade is in essence a general slackening of the money demand for commodities, and an expansion of trade is a general augmentation of the money demand for commodities." (Hawtrey (1913), p. 272)

The core of his theory is the assumption that the credit is inherently unstable and tends to fluctuate widely, bringing the rest of the economy together. In the Hawtrey model, the monetary policy can be used as a counter-cyclical instrument to mitigate or cancel the effects of crises and fluctuations.

2.5.2 The real economic theory of cyclical fluctuations of Robertson

Despite the fact that Robertson and Hawtrey were contemporaneous and they were associated with the same school of thought, he developed a completely opposite theory from the one described in the previous subsection. The Robertson theory (Robertson , 1915, 1948) assumes that cyclical fluctuations are mainly, or even exclusively, attributed to real factors. Innovations can create jumps in the economic growth and they push entrepreneurs to launch new investment plans to profit of new technologies. Since the investment plans of entrepreneurs are not known among them, this will lead to a situation of over-investment which then move the economy from its equilibrium position because the amounts of saving is fixed in a given period. In his view, economic fluctuations are an unavoidable component of economic growth related to technological development and innovation. Following Robertson's theory and in opposition to Hawtrey's one, monetary policy cannot have a counter-cyclical role but a pro-cyclical one so that an active monetary policy will likely increase the effects of a crisis rather than mitigating them.

2.5.3 The post-great depression theories

After the great depression of 1929, there was a great impulse in studying the nature, the causes and the consequences of cyclical fluctuations. Despite the fact that such studies did not lead to a unique theoretical view, following Fayolle (2003), we prefer to group them since they are characterised by the common factor constituted by the 1929 great depression.

The Frisch theory

For the first time, Frisch (1933) hypothesise that cyclical movements are generated by the accumulation over the time of exogenous shocks which tend to move the economy out from its equilibrium position. This theory was mainly based on the works of Slutsky which can be considered one of the pioneers of modern time series analysis.

As described in Fayolle (2003), "the impulses or shocks ("*the exterior impulse" or "erratic shocks'") that give rise to an economic cycle have a stochastic character and are therefore rather irregular; the propagation mechanism ("the intrinsic structure of the swinging system") transforms the sequence of such impulses into a succession of cycles whose length depends on the parameters of that mechanism, but whose amplitude is influenced by the strength of the impulses"*. The propagation mechanism is mainly based on the acceleration principle linking the level of orders and the production of investment goods to the variation in the production of the consumer goods.

The Frisch theory combines a stochastic component represented by the exogenous impulses affecting the economic equilibrium with a more deterministic propagation method which generate the cycles. The interaction between stochastic and deterministic factors can lead to a complex situation which can generate cycles of various periodicities, not necessarily synchronised among them.

The Haberler theory

Haberler (1946) was conducting a detailed critical analysis of various theories of cyclical movements. His conclusion was that none of them was fully acceptable and that, instead of following one given theory, an in-depth exploration of cyclical movements could lead to better results. Based on this assumption, he was elaborating a detailed table of causes and factors which characterise expansion and recession phases as well as peak and troughs of cycles. In his opinion, an integrated view of the financial and real sector as well as an accurate description of their interaction are the basis of a better understanding of cyclical fluctuations.

2.5.4 Schumpeter's Innovation Theory

Schumpeter's thought was trying to better identify the different phases on the cyclical movements and the main factors intervening on each of them. Schumpeter's view on cyclical movements is only a part of a more general economic theory on the economic development of a capitalistic system. In a normal situation, without any relevant change from the technological or innovation point of view, the economy is supposed to be in a stable equilibrium such as a Walrasian equilibrium. This equilibrium situation can be also dynamic on time but with fixed parameters which do not allow for any change in the relationship between various agents and markets. Entrepreneurs are the drivers of the capitalistic development. The deviation from the equilibrium has microeconomic origins. It is generated by the need of entrepreneurs to innovate and by the provision of credit to finance such innovations by banks (see Fayolle (2003) and Dal-Pont Legrand and Hagemann (2005)). An expansion phase is generated by innovations which mobilise resources. During this expansion, resources are redeployed in favour of innovating entrepreneurs. At the maturity of this process, new products will arrive on the market and the equilibrium is changed. The return back to the equilibrium or to a new equilibrium is then obtained via a recession phase where competing activities are eliminated. The process described can be considered as the simplest outline of cyclical fluctuations as an alternate of prosperity and recession phases. During such periods also the equilibrium will evolve along a trend which is endogenously determined by the sequence of innovations. This can be considered as a first example of the economic steady-state around which the economy fluctuates. The simple outline of cyclical movements can easily become more complex due to secondary effects induced by innovation which can generate a secondary cycle. In this context, it is possible that speculative bubbles occur with risks of serious depressions.

In Schumpeter's view, a full cycle is characterised by four well distinct phases: prosperity, recession, depression and recovery. This characterisation of cycles is not substantially different from the one proposed by Burns and Mitchell (1946) which will be described in section 2.7. Nevertheless, the two approaches, as we will see, are substantially different essentially because, in Schumpeter, the cycle is mainly viewed as a deviation from an equilibrium path, while in the Burns and Mitchell approach, the trend and cycle are considered together.

Schumpeter (1939) went beyond the assumption of a single cycle characterising economic fluctuations. Starting from the plurality and the heterogeneity of innovations he hypothesised a systematic approach to cyclical movements based on various cycles, each of them having distinguishing characteristics in terms of length, driving forces, etc. Originally, he was proposing a three cyclical scheme: short (Kitchin), medium (Juglar) and long (Kondratieff). In this representation he was ignoring the Kuznets cycle because he did not consider it as an acceptable one, since three Kuznets cycles were producing one Kondratieff. Nevertheless, we prefer here to present the scheme with four cycles as it is quite common nowadays. Table 2.1 will present the multi-cyclical scheme in a synthetic way, while the following paragraphs will analyse in some detail the four different kinds of cycles and provide a description of the multi-cyclical scheme as well as some concluding remarks.

Table 2.1: Overview: Multi-cyclical schemes

Name	Duration	Type
Kitchin	3 to 5 years	Inventory cycle
Juglar	7 to 11 years	Fixed-investment cycle
Kuznets	15 to 25 years	Infrastructural investment cycle or building cycle
Kondratiev	45 to 60 years	Long technological cycle

Inventory or Kitchin cycles

In 1923, Joseph Kitchin published an article in The Review of Economic Statistics (1923) called "Cycles and Trends in Economic Factors" (Kitchin (1923)). In this article he outlined his discovery of a 40 months cycle resulting from a study of US and UK statistics from 1890 to 1922. At the same time, a fellow Harvard professor — W.L. Crurn — found a cycle of 40 months in commercial paper rates in New York. A Kitchin

cycle is probably the best-known short-term cycle in the economy ever observed. A Kitchin cycle is tied to investment expenditures for inventory capital or consumer goods.

It is observed that inventory fluctuations are relatively more important than in earlier times because the other components of real output tend to be far less volatile. What causes the fluctuation in inventory cycle? It is a holding of inventories such as:

- to smooth production

- to produce more cost-effective lot sizes

- to buffer stock thereby preventing lost sales because of insufficient stock (a transaction motive)

- to take advantage of a lower prices (a speculative motive).

One could also say that it is based on a stocking/de-stocking cycle. Because consumption and production alternate, so do inventories. That is precisely why it is also called a "business cycle".

Fixed investment cycles or Juglar cycles

One of the first attempts to formulate an economic theory of cycles is due to the French economist Joseph Clement Juglar who studied the rise and fall in both interest rates and prices. Juglar (1862) argued that the credit cycle is closely related to the economic cycle. The link between the two cycles is represented by the speculative behaviour amplified by a collective dimension: the contagion effect (see Dal-Pont Legrand and Hagemann (2005)). The transmission mechanism described by Juglar starts from an excessive speculation which has two main consequences: generating financial bubbles and absorbing capitals and perturb investors. In such a context, some fixed good investments are either delayed or cancelled due to the fact that capitals are used more for speculative reasons instead of financing them. This process is well described in Dal-Pont Legrand and Hagemann (2005).

This cycle is attributed to investment expenditures on machinery, equipment and other similar capital goods. He observed a 7 to 11 year cycle. This cycle is characterised by a very negative view of Juglar on financial markets so that in his view, economic fluctuations are considered unavoidable. In the work of Schumpeter (1939) there was a great interest in the Juglar analysis but, while Juglar cycle was fully included in the Schumpeterian scheme, he mitigates quite a lot the Juglar position having a much less negative view on financial markets. It is important to observe at the end of this subsection that, among the cycles identified by Schumpeter, the Juglar one is probably the closest in terms of definition and characteristics to the modern definition of cycles.

The building cycle or Kuznets cycle

Russian born economist Simon Kuznets who finished his studies in America also found a cycle. Graduating at Columbia University he met Wesley C. Mitchell and the latter's work inspired Kuznets to study data on national income and capital formation. He identified a cycle of 16,5 to 18 years.

This cycle is both a long-term and short-term. The short-term fluctuation is tied to the credit markets. The long-term wave is primarily a function of demographics. The basic logic for the Kuznets cycle is as follows: During prosperous economic times the demand for labour increases, which puts upward pressure on wages. In turn, the improved economic environment causes an increase in new family formations, sparking the demand for new housing units. The dating of the Kuznets cycle seems more controversial.

The Building Cycle was constructed for understanding of the phases of the real estate cycle which is very important for investment edge timing. The Building Cycle is divided into four phases:

- development

- overbuilding

- adjustment

- acquisition.

The Building Cycle Phases for distinguishing particular phases are monitored a supply and demand cycle represented by home sales (demand side of the market) and housing starts (supply side of the market). Development-demand picks up and an increasing in housing starts follows. This phase is characterised by low vacancy rates and rising rents. This one reaches maturity after about 3 to 5 years. A signal of this turning point is the aggressive bidding up of land prices.

- **Overbuilding** - is the phase when housing starts consistently outpace home sales.

- **Adjustment** - is the phase when builders react to the declining demand and curtail housing starts.

- **Acquisition** - is the phase of the cycle when starts continue to decline while home sales are firm. The building activity is further reduced although vacancy rates have peaked and rent concessions have ceased.

However, criticism that a Kuznets cycle represented waves of population and migration didn't prevented him from receiving a Nobel Prize for his work.

Kondratieff cycles

Russian economist Nikolai Kondratieff worked in the Agricultural Academy and Business Research Institute in Moscow during the Stalin years. Kondratieff analysed the long wave cycles with duration of between 45 and 60 years. Kondratieff never set forth any theory for long wave. He only cited several empirical characteristics of his long wave. But his observations led to a theoretical explanation of the long wave by J. Schumpeter, W. W. Rostow, G. Ray, and J. van Duijn. All have developed the innovation theory of the long wave.

In the early 1920's he was committed to produce a study to show that capitalism would inevitably fail. His article "The long waves in economic life" published in The Review of Economic Statistics (1935) covered the major economies of the time: USA, Great Britain, Germany and France. A Kondratieff Cycle or Long Wave is a continuous repetitive pattern, also observed in 1913 by Dutchman Van Gelderen. He considered wholesale prices, interest rates, wages, foreign trade, industrial productivity and commodity prices from 1789. In 1926 Kondratieff demonstrated a cycle of economic activity roughly 54 years long (Kondratieff (1926)). This cycle or economic trend, from prosperity to depression, tends to repeat itself approximately every 54 years. The cycle shows repeated periods of ever increasing economic expansion and contraction. Kondratieff observed that from the despair of depression capitalism gave birth to another cycle of expansion mostly founded on a new technology. This new technology was only partially developed in the late stages of the previous cycle. This demonstrates the historic inherent ability of free people in a market environment (capitalism) to survive. His study in effect showed an inherent self-correcting mechanism perpetuating capitalism. This was of course unacceptable to his Communist masters. He was exiled to Siberia where he was quickly condemned to solitary confinement.

According to Simon Kuznets' article "Schumpeter's Business Cycles" published in The American Economic Review (Kuznets (1940)) a Kondratieff cycle consists of four distinct phases, or distinguishable, dramatic mood changes. These mood changes reflect and determine the actions of individuals involved in an economy. The mood of the up phase is one of accumulation, expansion and the desire for new product manufacture. The mood of the down phase is one of saturation and contraction. Dutch analyst and trader Koen Hoefgeest - visualised the subdivision of a Kondratieff wave as shown by the chart below.

Revival The first part of the up phase is associated with increasing economic activity, confidence, inflation, debt, prices and decreasing unemployment. Much effort is put into controlled growth. A booming stock market reflects confidence.

Figure 2.3: Schumpeter's Business Cycle
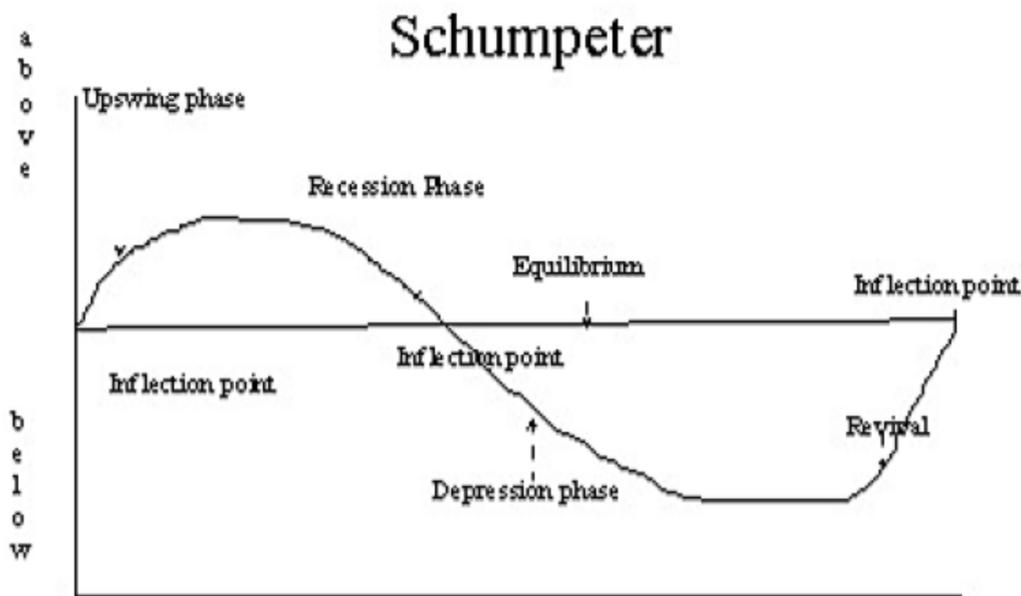
Recession The second part of the up phase is associated with ever increasing activity, debt, prices, inflation and confidence. At the end of this period inflation peaks just like prices. During summer an economy seems to get out of control. This is reflected in a faltering stock market.

Prosperity The first part of the down phase is associated with declining interest rates, prices, increasing debt, business activity and confidence. During this period, stock markets boom because of efficiency, worldwide price competition and globalisation.

Depression The second part of the down phase is associated with a debt-funded expansion (over production) which cannot sell its products to a debt saturated and product-loaded consumer. The massive debt implodes resulting in increasing contraction, declining prices and interest rates. When the debt problem is resolved a new birth of expansion is underway. Depressions are inevitable but healthy.

The ability of predicting the future course of the economy, the overall price level, the bond market and the stock market can be seen as an additional interesting feature of the Kondratieff framework.

In Schumpeter's view, the Kondratieff framework constitutes the most important prognostication tool (see 2.3). To understand how this tool can work, we consider the exercise proposed by Ivan Faustyn. He has plotted the US long bond yield and the constant dollar Dow Jones Industrials Average (price adjusted) in a Kondratieff framework. By applying Kondratieff properly one holds a long-term investment model: when what to buy and how long to hold it. Based on this exercise, the following Kondratieff phases were identified.

Revival (1949-1966) Out of the ruins of World War II 1949 gave birth to a new inflation/deflation cycle. The western economy was rebuild and by doing so people (re)gained confidence. This optimistic mood resulted in increased business activity and rising stock prices. Suddenly everybody realised that capitalism didn't fail. Governments and Central Bank didn't want the economy to grow too fast so they tried to control inflation and wages. The first part of the up phase peaked around 1966.

Recession (1966-1981) A long hot summer followed. Prices skyrocketed - helped by two OPEC shocks - and so did inflation. This changed the mood dramatically which was reflected by an invisible stock market crash (we have made the crash visible by measuring it in real terms or constant dollars). Too much money, debt and capacity had been created. Inflation (the up phase) peaked around 1981.

Prosperity (1981-1999) Inflation had caused lower demand so businesses turned their attention to efficiency: doing more with less resulting in lower prices. Global price competition increased and in order to survive

businesses had to cut costs. This changed the mood to optimism once again. The stock market took off like a space shuttle. Everlasting prosperity ahead. And by creating even more debt we can now say we are experiencing a credit card economy.

Depression (2000-2009) Winters are hard but very healthy. During such a period too much debt, money and capacity will be eliminated. Debt liquidation will change the mood and pave the way for a depression. This will be reflected in a massive drop in stock prices (90% will be lost).

A detailed analysis of the Kondratieff cycle can be found in Carry (2003). Most empirical evidence does not support the long wave lasting about 50 years and resulting from a clustering of innovation. Even if a 9-year centred moving average of the data is calculated, as Kondratieff suggested, it is difficult to see a pronounced long wave in this data over this period. It is necessary also to discriminate the influence of the periodicity of the moving averages estimators from the periodicity which really exists (Slutsky-Yule effect). In addition to these considerations, it is clear that the identification of a long cycle such as the Kondratieff one will require availability of statistical data over a very long period which is not very easy. Furthermore, even if such data exists, their quality and especially their reliability cannot be considered very high. Finally, working over a very long time period implies the presence of a large number of structural breaks due to technological progress, catastrophic events such as wars, changes in the geographical definition of a country, etc.

The presence of such breaks will further complicate the analysis. Under this situation it is quite clear that many economists and researchers do not support the existence of the Kondratieff cycles. Granger and Hatanaka (1964) also said that it is almost impossible to distinguish between going on a growing trend or on an expansion phase of the Kondratieff cycle due to the limited number of available observations. Despite this scepticism, the Kondratieff cycle remains a fascinating idea and also a very interesting theoretical exercise. Also for those reasons, in the last years, some studies have tried to revitalise the Kondratieff cycle and to find empirical evidence of its existence. Among them, probably the most interesting one has been published by Korotayev and Tsirel (2010). One of the main conclusions of this paper is that there is statistically significant evidence, based on spectral methods, of the presence of the Kondratieff cycle on the world GDP from 1870 to 2007.

Nevertheless, the possibility of identifying the Kondratieff cycle and then to validate the assumptions underlying its existence remains still an open problem without solution.

The multi-cyclical scheme

In the scheme proposed by Schumpeter, the Kondratieff cycle plays the role of the framework. Inside the framework, shorter minor cycles can then occur. Each cycle has well distinguished origins because the kind of innovations generating it is clearly distinguished by other innovations. In this way, various cycles are mainly uncorrelated but they are linked by some cross-cyclical constraints on their length. More in detail, one Kondratieff cycle is equal to approximately six Juglar cycles and one Juglar cycle is equal to 2-3 Kitchin cycles. A very comprehensive explanation of the Schumpeter view of cyclical movements is provided by Geiger (2012). He also provides an interesting graphical representation of the multi-cyclical scheme based on the original figure in Schumpeter (1939). This graphical scheme is presented in figure 2.4.

Some concluding remarks

The Schumpeter contribution to the theoretical definition and explanation of economic cycles is, without any doubt, the most comprehensive and interesting work of the pre-Second World War period. Schumpeter's contribution is still very relevant and it has influenced most of the theories and studies on cyclical movements. Apart the development of its innovative multi-cyclical scheme, we want to emphasise the fact that in his view, cyclical fluctuations do not have necessarily a negative connotation. They are generated by the innovation process and the innovation is the main driver of economic development. Fluctuations move the economy from an equilibrium position to a next one which is likely better than the previous, due to the fact that innovations have developed the economy. This approach is in contradiction with most of the contemporaneous authors

Figure 2.4: Replication of Schumpeter's original superposition graph from the Business Cycles (1939). The continuous line displays the sum of the three dashed cyclical movements added up onto one another.
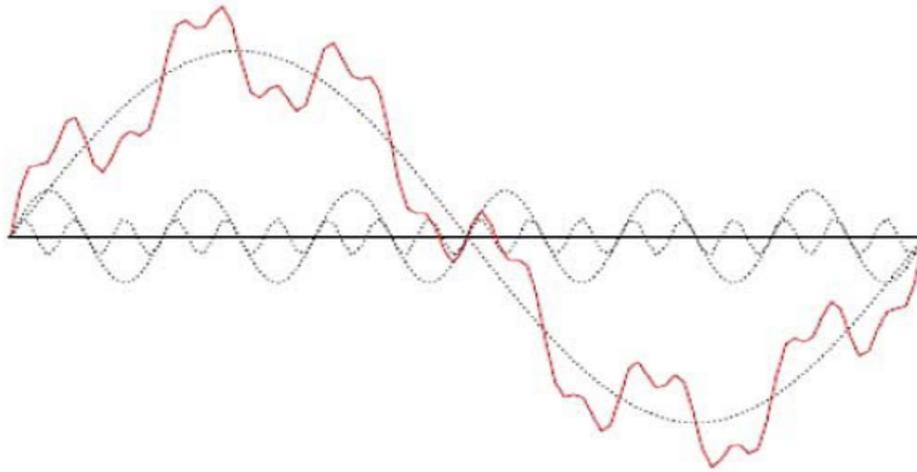
for having a negative interpretation of economic fluctuations. In this context, also the Kondratieff cycle was receiving a constructive interpretation, despite the fact that originally the Kondratieff cycle was considered the way to demonstrate the unavoidable end of the capitalistic economy. Even this assumption was never demonstrated by Kondratieff.

2.5.5 Psychological theory of cyclical fluctuations

The idea that psychological factors can move the economy from an equilibrium position and start cyclical movements has been present in the economic literature since its very beginning. This idea was not necessarily related to a well-defined theory (classical, Keynesian, etc.) but it was largely spread across them. Looking at the economic literature, it is then possible to find a number of economists, who have explored mass psychology and its consequences for economic behaviour.

The main assumption of the psychological theory is that individuals are strongly influenced by the beliefs of the group or groups to which they belong. There are times when the general mood is optimistic, and others when it is pessimistic. In this subsection, we briefly analyse this theory by focusing mainly into two contributions: the first one from Stuart Mill (1826) and the second one from Pigou (1927).

The contribution of Mill

Even if Mill is usually considered, together with Smith and Ricardo, the major author of the classical macroeconomic theory, he is probably the only one, at the beginning of the 19th century, paying attention to psychological aspects as causes of cyclical fluctuations. Mill's thought on cyclical fluctuations is described in his paper "Paper currency and commercial distress" (Mill (1826)) and a very detailed analysis of his theory is presented in Forget (1990).

For Mill, the key, but quite simplistic, distinction is between rash speculators and professional gamblers. Rash speculators are attracted by a market when the price increases. This attitude is motivated by the fact that rash speculators do not know the main drivers of supply and demand and this makes the main difference with professional gamblers. Consequently, due to this different behaviour, any shock leading to price increases is a potential impulse to speculation and to instability. Beside this, there are also some structural reasons generating fluctuations such as the opening of new markets and the so called calm situations. In particular, in a calm period professional gamblers tend to exploit profit accumulating capital, but on the same time the

interest rate on capital decreases so that some professional gamblers tend to become speculators. But the speculation can be unsuccessful if the gamblers stop buying from the speculators which creates panic among them. The panic can move the actual price down to the equilibrium price, mainly because, in a panic context, banks stop to supply credits to them.

In addition, Mill, anticipating Pigou, also sees in the process of expectations formation of economic agents another potential source of economic fluctuations.

The contribution of Pigou

An English economist, Arthur C. Pigou, in his *Industrial Fluctuations* (Pigou (1927)), put forward a theory of "non-compensated errors". He pointed out that if individuals behave in a completely autonomous way their errors in expectations would tend to offset each other. But if they imitate each other, their errors will accumulate until they acquire a global magnitude that may have powerful economic effects. This follow-the-crowd tendency obviously operates as a factor in the ups and downs of the stock exchanges, financial booms and crashes, and the behaviour of investors. One could argue, however, that the psychological factors are not enough to explain economic fluctuations. The scepticism on this theory is based on the hypothesis that moods of optimism and pessimism should themselves rely on economic factors.

Nevertheless, following Pigou we can assume that psychological factors can play a role, even if not an exclusive one, in changing the economic situation. An interesting review of the Pigou theory, also in the light of contemporaneous economic school of thought, is provided by Collard (1996). At the same time, starting by the non-technical exposition provided by Pigou (1927), Beaudry and Portier (2004) provide an interesting formalisation of the Pigou model. Following Collard (1996), the Pigou business cycle theory can be synthesised as follows:

"Fluctuations (irregular cycles) are driven by variations in the profit expectations of business people. These, in turn, are set off by 'impulses' which may be 'real', 'psychological' or 'monetary' in nature. Once the initiating impulse has made itself felt, it may be sustained by any or all of the three. Consequent fluctuations in employment will be greater the greater the degree of price or wage rigidity and are (because of externalities) probably larger than is socially desirable."

Just as a concluding remark, we consider that it is important to notice that Pigou, by accounting the role of future uncertainty in its mechanisms of generating expectation, was anticipating the Keynesian thought which bases its view of economic instability on the uncertainty characterising future capital profits.

2.6 The Keynesian theory

The publication of the book *The General Theory of Employment, Interest and Money* by J.M. Keynes (1936) constituted a breakthrough for the economic theory. In this book, Keynes strongly contradicts the main assumption of classical economy and describes a completely new way of economic functioning which has influenced economic thought in a permanent way.

The Keynesian critique of the classical self- correcting mechanism is presented in two categories:

- the failure of demand to adjust because of monetary impotence (the failure of real GDP to respond to an increase in the real money supply or a fall in the real interest rate), and

- the failure of supply to adjust as a result of rigid wages (the failure of the nominal wage rate to adjust by the amount needed to maintain equilibrium in the labour market.

Below we are going into some details concerning the Keynesian criticism to the demand and supply adjustment respectively.

A) Demand adjustment

The classical model, with perfectly flexible prices, relies on two important mechanisms by which a deflation (a fall in the level of prices,) leads to a stimulation of output to its natural level:

- a deflation has to lower interest rates through increasing real balances , and

- interest rates must be lowered by enough to stimulate the planned autonomous spending and aggregate demand necessary to bring the output level back to its natural level.

Keynes' objection to the first channel is the possibility of a liquidity trap, in which an extremely low interest rate causes people to hold any additional money instead of purchasing interest-bearing assets. The liquidity trap corresponds to a perfectly flat money demand schedule and LM curve so that interest rates — and thus output — will not respond to the increase in real money supply resulting from the deflation.

Similarly, Keynes' objection to the second channel is the possibility that planned autonomous expenditures are very or totally unresponsive to changes in the interest rate. This implies a very steep or vertical IS curve and AD curve, so that output will not respond to deflation. Because both of these demand-side problems are the result of the failure of flexible prices to influence real output through the real money supply, they are called the problem of monetary or deflation impotence.

In response to Keynes' criticisms, the classicals responded with the argument of the Pigou or real balance effect (the direct stimulus to aggregate demand caused by an increase in the real money supply; does not require a fall in the interest rate), where the increase in real money balances caused by a deflation stimulates autonomous spending, and thus the IS curve directly.

The force of the real balance effect is countered, however, by the possibility of the destabilising expectations effect (the decline in aggregate demand caused by the postponement of purchases when consumers expect prices to fall in the future) and redistributionary effect (the fall in aggregate demand caused by the effect of falling prices in redistributing income from high-spending debtors to low-spending savers) of falling prices.

B) Supply adjustment

Keynes' second critique of the classicals is that nominal wage rigidity will imply a failure of the aggregate supply curve to adjust the economy to the long-run equilibrium level. Graphs of the labour market and the output market show the implications of the Keynesian rigid-wage assumption in the context of the derivation of the SAS curve.

The situation of a fall in aggregate demand along a given aggregate-supply curve, and the subsequent rise in the real wage rate in the labour market, is perfectly analogous to the short-run equilibrium point resulting from an increase in aggregate demand, but the essential difference in the Keynesian model is that nominal wage rigidity keeps workers off their labour-supply curve, preventing the adjustment of the actual wage to its equilibrium market-clearing value, and thus preventing the shifts in the SAS curve necessary to return the economy to its natural level of output (Y_N). The result is persistent unemployment.

Although the Keynesian model is able to explain the persistence of unemployment from the excess supply of labour that arises from nominal wage rigidity, its drawbacks include:

- a failure to explain why or how the nominal wage remains rigid, and

- the requirement that real wages move countercyclically.

In particular, Keynes states that markets are often imperfect which prevents a self-adjustment of the economy unless there are interventions represented by active monetary and fiscal policy. The *General Theory* describes a sequence of events which can lead to expansion and recessions but it does not have any clear statement about cyclical fluctuations. There is not an automatic sequence of cycles as in Schumpeter but fluctuations can occur on the basis of the degree of uncertainty of expectations and on expectations divergence across economic agents. Consequently we can argue that in the purely Keynesian model there are no regular cyclical fluctuations.

Nevertheless, it is possible to identify some relations which can lead the economy to fluctuate. This is the case of the divergence between the aggregate demand and supply which generates fluctuations and which move the economy to a short run equilibrium position at levels that are different from the full employment compatible ones.

More in detail, fluctuations in the interest rate equated savings and investment. Any divergence between savings and investment would affect the interest rate so as to re-establish equality between savings and investment. Keynes emphasized the role of effective demand-the demand of consumers for consumer's goods, firms for investment goods, government for public goods.

Keynes asserted the consumption function, representing the functional relationship between income changes and consumption changes, was one of the most stable relationships. Investments were determined by the profit expectations of the entrepreneurs and the cost of investments. Investments could be affected by changes in the interest rate. Cyclical fluctuations are largely the result of instability in private investment. In the Keynesian model, the marginal efficiency of investment reflects entrepreneurs current profit expectations.

Main causes of the Keynesian crisis can be:

- complete collapse in the marginal efficiency of investment;

- little sensitivity to changing interest rates in the economy;

- profit expectations so low as to preclude investment in the eyes of the entrepreneur, even at low interest rates.

According to Keynes, an "underemployment equilibrium" is only possible if wages and prices are 'sticky'. If prices (and wages) are flexible then there is always a market price, which will equate demand and supply. The recovery from the crisis is not an automatic process, but it requires interventions of public authorities in the form of monetary and fiscal policies. Monetary policy can encourage activities in private sector even when entrepreneurs' profit expectations could be so low that any interest rate seems too high. Fiscal policy can stimulate economic activity by direct government expenditures, mainly in infrastructures, and taxes.

Some authors viewed in the absence of an explicit mechanism generating economic fluctuations one of the limits of the Keynesian theory. In the Keynes work there is the emphasis on the role of expectations and on the fact that there is not a unique way to exit from a depression condition. Those ideas were then formalised and systematised in Hicks (1939). In the Keynesian theory there is also a psychological mechanism as a potential source of economic instability. This is motivated by the fact that investments cannot be explained by rational choices. Keynes argues that the driving force of investment is animal spirits; defined as "spontaneous urges to action rather than inaction" (Matthews (1984), Keynes (1936)).

The possibility for the Keynesian model to experience endogenous fluctuations was developed by some authors who were trying to complete and extend the Keynesian theory in several directions to better account for economic instability and fluctuations. In the following subsections we will shortly present some of them, focusing on those developed more or less at the same period of the publication of the *General Theory*.

2.6.1 The Harrod-Domar model

This model, mainly oriented to describe long term economic growth, was independently developed by Harrod (1939) and Domar (1946). The model explains the relative changes in economic growth (i.e. growth rate of GDP) in terms of the level of saving and productivity of capital (the ratio between capita and output). The model shows that there is no natural reason for an economy to have balanced growth. Three different kinds of growth can be defined within this model: warranted growth, actual growth and natural rate of growth. Warranted growth rate is the rate of growth at which the economy does not expand indefinitely or goes into recession. Under warranted growth, the whole amount of saving is transformed into investment. The actual

growth is the observed growth of the economy. The natural rate of growth is the rate of economic growth which guarantees the full employment.

2.6.2 Samuelson's oscillator model

The so called Samuelson's oscillator model (Samuelson (1939a), Samuelson (1939b) can be viewed as a dynamic extension of the basic Keynesian model. In detail, this model contains two behavioural equations: one for the consumption which is the dynamic version of Keynesian consumption function and one for the investment which are endogenised into the model via the so called acceleration mechanism. In this model the only exogenous variable remains government expenditure. The accelerator equation is expressed in such a way that the investments depend on the changes in consumption. An in-depth analysis of this model can be found in Heertje and Heemeijer (2002). Beside the dynamic specification of the model, the real innovation with respect to the Keynesian theory is constituted by the specification of the investment function via the accelerator model, instead of having exogenous investments as in the original Keynesian theory. The model is a second difference system which generates cyclical movements in the GDP for given values of the parameters. In the Samuelson model the amplitude of cyclical movements depends on the level of the investments, under the assumption that the investments determines the level of aggregate output (multiplier), and are determined by aggregate demand (accelerator).Extensions of the Samuelson model have been proposed by Kuznets (1940) and Kaldor (1954), Marrama (1946), and Goodwin (1951, 1967). A detailed comparison of the various accelerator mechanisms is provided in Pasinetti (1960).

2.7 The empirical revolution of Burns and Mitchell

After the poor performance of the Harvard economic barometers and their variants during the Great Depression of 1929, empirical studies aiming to analyse and forecast cyclical movements, knew a quite obscure period without relevant contributions. The end of this period has been represented by the publication of the seminal work of Mitchell and Burns (1938) and Burns and Mitchell (1946). In Burns and Mitchell (1946), they defined business cycle as:

"Business cycles are a type of fluctuation found in aggregate economic activity of nations that organise their work mainly in business enterprises: a cycle consist of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions and revivals which merge into expansion phase of the next cycle; the sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitude approximating their own."

Figure 2.5 describes a typical business cycle fluctuation underlying the various phases in which it can be subdivided according to Burns and Mitchell.

Since then the National Bureau of Economic Research (NBER) business cycle program, the U.S. Department of Commerce Bureau of Economic Analysis, and more recently The Conference Board have relied on the indicators approach that was based on this definition. The chronologies of business cycles developed based on this approach and the classifications of economic indicators into leading, coincident, and lagging indicators have been useful ingredients in the measurement and analysis of business cycles over the years in many countries.

After more than sixty years that definition is still frequently quoted. The Burns and Mitchell definition makes clear that the "business cycle" refers to an aggregate quantity representing all sectors in the economy. Hence, the practice at the NBER has been to use information from several sectors of the economy to decide which dates constitute the turning points of the cycle.

Figure 2.5: NBER representation of business cycle according to Burns and Mitchell's definition
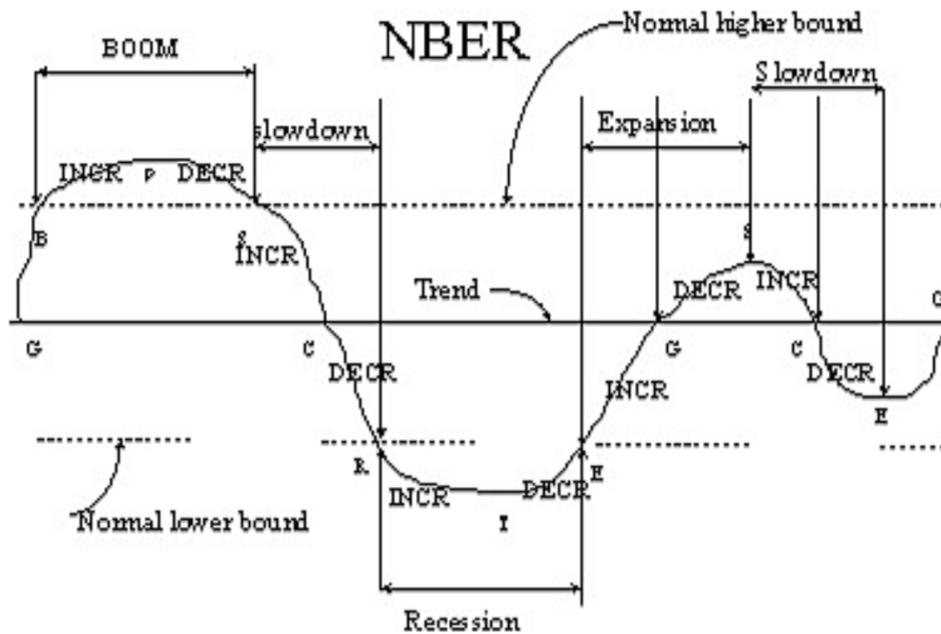
Implicit in the Burns and Mitchell's definition and in NBER methodology is an important assumption: that there exists an unobserved, common to all sectors, state variable which we call the "business cycle". Although some series may lead or lag the other they all share a relationship with the common variable, the business cycle which is unobserved. This means that, the GDP cannot be considered as the reference variable for business cycle analysis and that the reference variable is latent or unobserved. This latent variable can be expressed or approximated by a weighted average of a number of statistical indicators

As emphasised by Diebold and Rudebusch (1996), the two features of Burns and Mitchell definition, namely distinct expansion and contraction phases or regimes on the one hand, and common movements on the other, have given rise to two distinct but related approaches to business cycles modelling: the first one focusing on the asymmetric behaviour in various cyclical phases implying the existence of different regimes and the second one, mainly looking at communalities and common movements between economic variables.

Studies which pursue the regime approach include Neftci (1992), Diebold and Rudebusch (1989), Artis et al. (1995) and perhaps most prominently, Hamilton (1989). The simplest form has two regimes, corresponding to expansions and contractions, with the characteristics of economic activity assumed to be dependent on the regime. Using observed information, optimal (probabilistic) estimates of current or future regime can then be obtained. Whereas the studies of Neftci, Diebold and Rudebusch and Artis et al. (1995) consider leading indicators in the absence of an inherent regime-switching mechanism, Hamilton develops the approach by including a Markov process that governs the switch between regimes. Many extensions have been made to the Hamilton's model, including Filardo (1994) and Hamilton and Perez-Quiros (1996) who allow leading indicator information to be used to predict the future regime.

In the common movement tradition, a continuous business cycle indicator is usually estimated or extracted from observed values of economic variables. In contrast to the regime approach, this extraction takes place without assuming that the regime itself dictates the underlying characteristics. A good recent example is the dynamic index model of Stock and Watson (1989, 1991) and Stock and Watson (1991) which is based on earlier work of Sargent and Sims (1977).

2.7.1 Measuring the business cycle: the NBER approach

The NBER defines a recession as "a recurring period of decline in total output, income, employment, and trade, usually lasting from six months to a year, and marked by widespread contractions in many sectors of the economy."

Recessions are, therefore, macroeconomic in nature. A severe decline in an important industry or sector of the economy may involve great hardships for the workers and firms in that industry, but a recession is more than that. It is a period in which many sectors of the economy experience declines. Recessions are sometimes said to occur if total output declines for two consecutive quarters. However, this is not the formal definition used by the NBER because it is regarded as a too narrow a way to define an economy-wide recession.

Business cycle peaks and troughs cannot be identified immediately (i.e. in real time) when they occur for two reasons. First, recessions and expansions are, by definition, recurring periods of either decline or growth. One quarter of declining GDP would not necessarily indicate that the economy had entered a recession, just as one quarter of positive growth need not signal that a recession had ended. The recession of 1981-82 provides a good example. Real GDP declined from the third quarter of 1981 to the fourth quarter, and then again from the fourth quarter to the first quarter of 1982. It then grew in the second quarter of 1982. The recession was not over, however, as GDP again declined in the third quarter of 1982. Only beginning with the fourth quarter did real output begin a sustained period of growth.

Second, the information that is needed to determine whether the economy has entered a recession or moved into an expansion phase is only available with a time lag. Delays in data collection and revisions in the preliminary estimates of economic activity mean the NBER must wait some time before a clear picture of the economy's behaviour is available. For example, it was not until December 1992 that the NBER announced that the trough ending the last recession had occurred in March 1991, a delay of 20 months.

An extension of the business cycle definition is the concept of growth cycles. During periods of high growth trends, business cycle contractions and revivals (or turning points) are limited. Economic activity exhibits fluctuations during such periods but these are usually obscured by the growth trends. In these cases, growth cycles can be identified as cycles in deviations from a long term trend. This modified approach was first applied by Mintz (1969) and later by Klein and Moore (1985). These researchers identified growth cycles (or cycles in deviations from trend) in the post-World War II European economies which also exhibited strong growth trends and few business cycle recessions. Klein and Moore (1985) showed that the typical classification of measures of different types of economic activity into leading, coincident, and lagging with respect to business cycles also applied to growth cycles. In their development of the LEI Adams et al. (2010) also consider evidence based on growth cycles. Growth cycle analysis requires the estimation of long term trends accurately.

2.7.2 Composite indexes

Within the NBER approach, composite indexes serve as handy summary measures of the current behaviour of the cyclical indicators and of the state of the economy. Because they are averages, they tend to smooth out some of the volatility of individual series. Use of composite indexes is consistent with the traditional view of the business cycle developed by Burns and Mitchell. In particular, composite indexes can reveal common turning point patterns in a set of economic data in a clearer and more convincing manner than the behaviour of any individual component.

Still, the composite indexes have limitations. They can retain some erratic movements, especially in data for the latest months. Sometimes these fluctuations are removed by later data revisions and sometimes not, which adds another complication to the historical and real-time appraisal of the cyclical indicators.

Even after data revisions, the turning points of the composite indexes exhibit a certain degree of timing irregularity. Although it is reasonable to expect the composite leading index to lead the business cycle, and the composite lagging index to lag, they do so with considerable variability. It has always proven too much

to expect those indexes to lead and lag by consistent intervals. On occasion even the turning points in the coincident index, the most regular of the three composites, miss the official peak and trough months of the business cycle (as determined by the National Bureau of Economic Research), but it is almost always within roughly three months of the official business cycle dates.

2.7.3 Calculating the composite indexes

The composite indexes are constructed as equally weighted averages of the components' symmetric monthly changes. Components monthly growth rates are first volatility adjusted using inverse standard deviations of the monthly symmetric changes in the components (the inverse standard deviations are further normalized to sum to one). Symmetric monthly changes are essentially growth rates, equivalent to taking log differences of the variables. The monthly growth rate of the index obtained in the previous step is cumulated to obtain levels of the index and this is then re-based to a fixed base year so that the average of the index values in the base year equal 100.

A detailed presentation of the methodology used for the construction of composite indexes according to the NBER and Conference Board approach can be found in Ozyildirim (Chapter 8) while some additional considerations and examples are presented in Carriero, Marcellino and Mazzi (Chapter 3).

2.7.4 Diffusion indexes

Diffusion indexes, which measure the proportion of a set of indicators that are rising, provide another source of useful information about the business cycle. They tell us how widespread a particular business cycle movement (expansion or contraction) has become.

The Business Cycle Indicators database (maintained by The Conference Board) includes diffusion indexes over two different time spans, one month and six months, for the components of the leading, coincident, and lagging indexes and for employment in 356 industries. The one-month span indexes tend to be erratic; signals from six-month diffusion indexes are much more reliable. Industrial production indexes published by the Federal Reserve Board and the Employment Situation Release from the Census Bureau that report payroll employment figures for the United States also report diffusion indexes. For example, the employment diffusion indexes measure the proportion of industries with increasing employment. The composite and diffusion indexes are not redundant even though both indexes are based on the same set of data. On occasion, they move in different directions. The composite index differentiates between small and large overall movements in the component series, while the diffusion index measures the breadth of a general movement, not the size. The difference is often very useful when attempting to either confirm or predict cyclical turning points.

2.7.5 Forecasting recessions

Followers of the cyclical indicators are keenly aware of problems that arise when interpreting the leading index in real time. The long-standing rule of thumb that a three-month consecutive decline signals a recession has given at least two false signals during the expansion from November 2001 to December 2007. However, few economists actually use such an inflexible rule. Most require a significant downward movement in the index-of at least 1 percent-and declines in the majority of the component series.

One reason it is imprudent to forecast recessions using a simple, inflexible rule is consistent with that recessions are not dated using a simple, inflexible rule. The U.S. economy is continually evolving and is far too complex to be summarised by one economic series. Official recession dates are determined from a multitude of indicators and there is no agreed upon formula for setting peak and trough dates.

"Why not replace all this agonising over a multiplicity of measures with a simple formula - say, define a recession as two consecutive quarters of decline in GNP? Any single measure is sure to encounter special problems just when they matter the most. ... We plan to stick with examining all of the data we can and making an informed judgment." (Robert Hall, Chair of NBER Business Cycle Dating Committee)

Forecasting a peak, which is usually considered more important, than forecasting a trough, has to be considered an even harder task than determining one after the fact.

Even though the composite leading index has flaws, and is not 100 percent reliable, it is more reliable than its individual components used alone and it can be used along with the corresponding diffusion index to give useful signals about the likely direction of the economy. Historical analysis shows that a negative growth rate over a six-month period of about 4 to 5 percent for the leading index and declines in at least half of the components (i.e., the six-month diffusion index falling below 50 percent) is a reasonable criteria for a recession warning.

It should also be recognised that a "false signal" from either the leading index or an alternative set of cyclical indicators is not necessarily a flaw–sometimes they are quite insightful. In 1966 and 1984, for instance, the leading index turned down significantly but a recession did not follow. Because economic growth weakened slightly thereafter, many economists believe that the index warned appropriately that the risk of a recession had increased. It is as though the leading index spotted conditions that often lead to a tropical storm, but turned out to be nothing more than a rain shower.

2.8 Other definitions of cycles

Despite the fact that the business cycle based on Burns and Mitchell's definition has been and still is the core of contemporaneous analysis of macroeconomic cyclical movements, since the second half of the 20th century, several researchers started to investigate alternative ways of defining cycles. They were not really aiming to contradict the existence and usefulness of business cycle but more to find other perspectives which could complete and integrate the existing one with the objective of obtaining a better picture of the economic situation.

The first fact motivating such investigations was that, when economies are characterised by a fast and stable growth, it is unlikely that business cycle movements can be statistically detected mainly because the trend is playing a dominant role with respect to other components. Nevertheless, this situation does not mean that fluctuations are absent but that they are not easily detected since they are hidden by the trend. In such a case, removing the trend can help in making underlying fluctuations more visible. Obviously, also in period of strong growth, it is not excluded that business cycle fluctuations happen but they occur very rarely so that, concentrating the attention on the business cycle only can be useless in periods between two peaks. This situation was, for the first time, evidenced by Mintz (1969) followed by Klein and Moore (1985). The genesis and rationale of Mintz work is described in Qin (2010). Mintz, experiencing problems in dating German business cycles at the NBER following the Burns and Mitchell definition, decided to analyse German cycles once the trend was removed. She explored two alternatives: the first one based on detrending and the second one based on the use of monthly growth rate. Since growth rate data were too much irregular, she decided then to concentrate her attention on detrended data.

Later on, in the terminology commonly used by researchers and experts in the field of economic fluctuations, the Mitchell and Burns (1938) definition started to be referred as classical cycle or classical business cycles, where the analysis based on detrended data was referred as growth cycle or deviation cycle. The analysis based on growth rate was referred as growth rate cycle or acceleration cycle. The usefulness of such a new definition of cycle was further demonstrated by the so-called "great moderation" period where business cycle fluctuations were almost absent over a quite long period. The following subsection will briefly describe the main characteristics of the growth cycle and acceleration cycle while the last one will introduce a unified

approach. An in depth analysis of alternative definition of cycles is provided by Zarnowitz and Ozyildirim (2000).

2.8.1 Growth cycle

Beside the reasons mentioned above and related to the few number of business cycle movements in presence of strong economic growth, the need of analysing separately the economic growth and the cycle also justifies the use of growth cycle data. It is useful to notice as growth cycle has also strong links with economic theory, especially the one on potential output and output gap, based on the Okun's law, Okun (1962). If we assume that a reasonable estimate of the trend can be seen as a statistical proxy of the potential output, then the growth cycle can be viewed as a statistical estimate of the output gap.

Furthermore, the decomposition into trend and growth cycle has also relevant similarities with the distinction between permanent and transitory component or permanent and transitory shocks quite common in the neo-classical literature but also in some post-Keynesian authors. The crucial point, when identifying the growth cycle, is constituted by the way in which data are detrended. The first detrending method called phase average trend (PAT) was developed by Boschan and Ebanks (1978). Since then, a variety of detrending methods has been proposed by several authors following either non-parametric or parametric approaches. A review of the most commonly used detrending filters is presented in Mazzi, Mitchel and Ozyildirim (Chapter 9).

Growth cycle estimates are usually more symmetric than business cycle ones and they are commonly characterised by two phases usually referred as slowdown and speed-up or recovery. Klein and Moore (1985) showed that the typical classification of measures of different types of economic activity into leading, coincident, and lagging with respect to business cycles also applied to growth cycles. In their development of the LEI Adams et al. (2010) also consider evidence based on growth cycles. Growth cycle analysis requires the estimation of long term trends accurately. Business cycle upper turning points usually anticipate, or at least coincide with business cycle ones, while troughs of growth cycle usually follow those of business cycle. It is also useful to notice that growth cycle fluctuations will not necessarily determine business cycle ones, so that the number of growth cycles in a given period usually exceeds the number of business cycles.

Despite this theoretical and empirical interest, working with growth cycle data presents some relevant drawbacks of which users should be well aware. The first one is represented by the fact that growth cycle estimates are conditional to the chosen detrending or trend estimate methods. See Mazzi, Mitchel and Ozyildirim (Chapter 9) for more details. Figures 2.6 to 2.8 provide an example on how growth cycle estimates are sensitive to the detrending methods. The three graphs show alternative trend-cycle decompositions of the euro area GDP according to Hodrick and Prescott (1980), and Christiano and Fitzgerald (2003) filters, and to unobserved component model.

Figure 2.6: Euro area trend cycle decomposition of industrial production index: Hodrick and Prescott filter
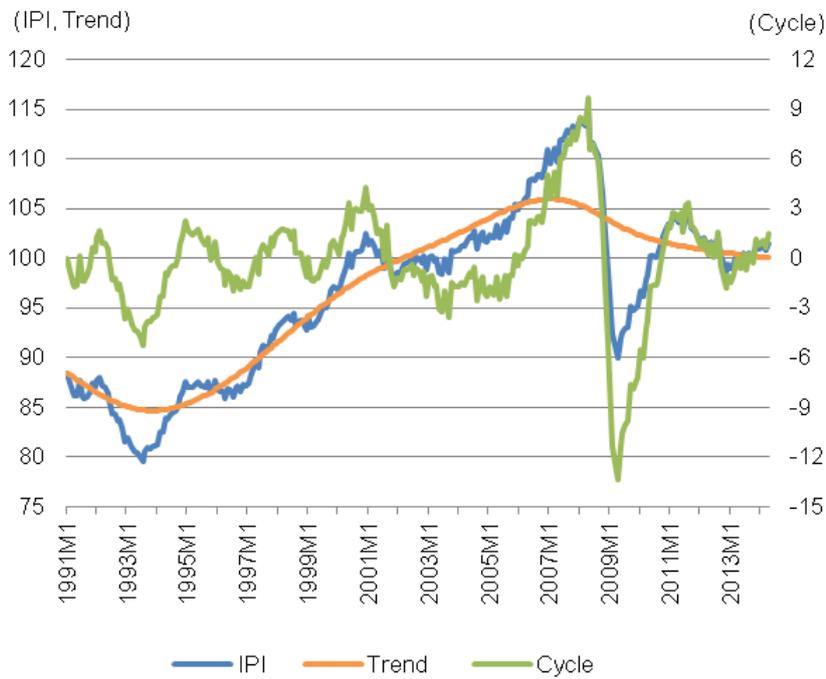
Figure 2.7: Euro area trend cycle decomposition of industrial production index: Cristiano and Fitzgerald filter
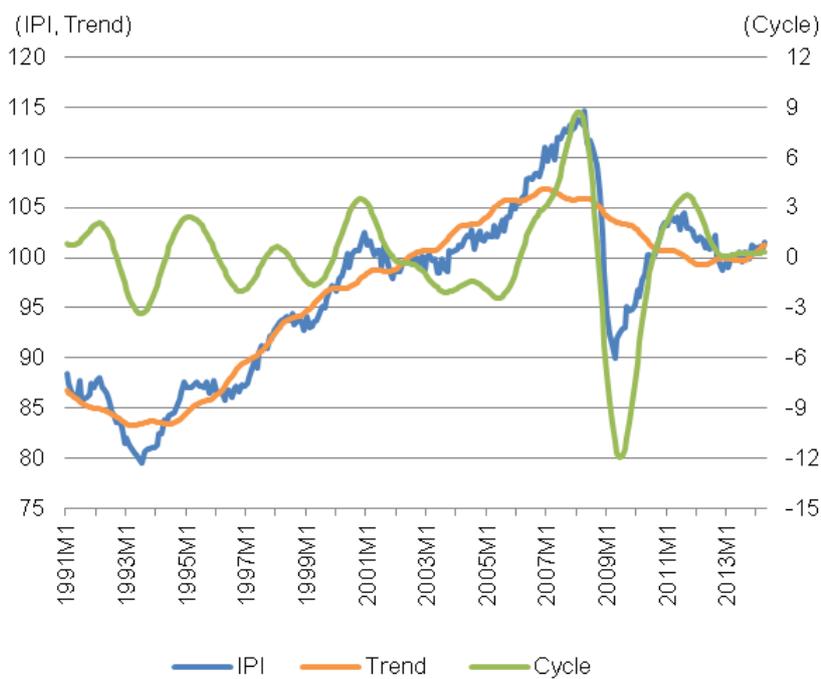
Figure 2.8: Euro area trend cycle decomposition of industrial production index: unobserved component model
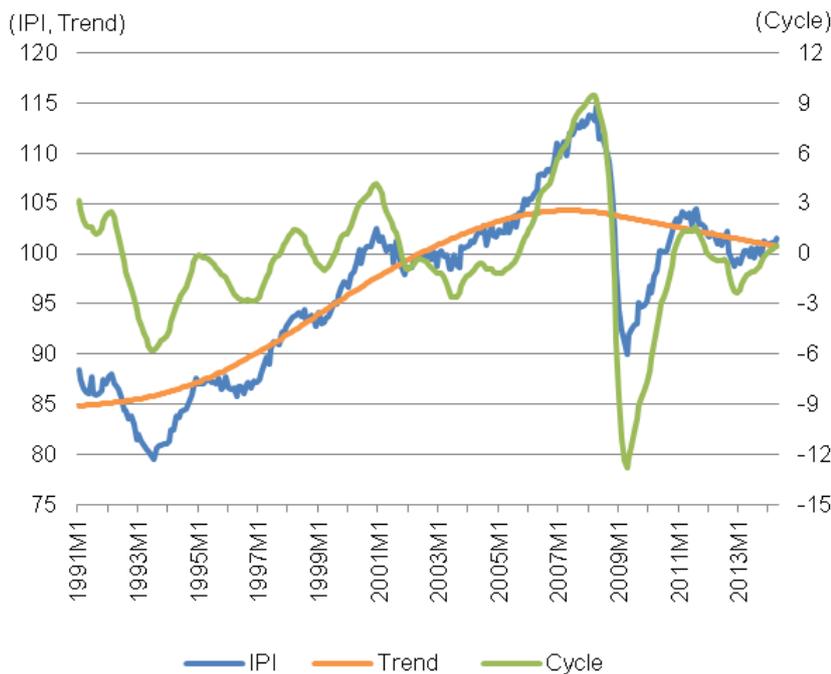
The second one is related to the fact that, especially when using linear filters such as moving average ones, current data can be either unavailable or not as reliable as the rest of the series which can compromise the real-time investigation. See Orphanides and Van Norden (1999), Orphanides and Van Norden (2002). Finally, it can be interesting here to recall that Central Banks attach a great importance to the availability of real-time output gap estimates since the output gap is considered a good predictor of future inflation.

2.8.2 Acceleration cycle

To overcome the drawbacks characterising the growth cycle, Friedman and Schwartz (1963) and Mintz (1969) were analysing the possibility of working on monthly growth rates under the hypothesis that the large part of trend effects could in this way be removed. Unfortunately, as noticed also by Mintz (1969), growth rate series are too erratic. Even when they are derived by series with well identifiable cyclical components, they tend to be dominated by the irregular components. The only way to make growth rate data useful is to smooth them, for example by using some filters or moving averages. This could introduce some phase shift with negative consequences for their usefulness. Nevertheless, especially in the last years, a certain interest has been devoted to the acceleration cycle. The main reason is that, under the hypothesis of a clear identification of its movement, the peaks in the acceleration cycle tend to anticipate those in the growth cycle by creating a very interesting warning system. The acceleration cycle has to be managed and considered with particular carefulness by the users, especially due to its erratic behaviour which could lead to false signals.

2.8.3 Unified framework

The scheme with three cycles described in this section – classical, growth and acceleration – could lead the reader to find similarities and analogies with the scheme proposed by Schumpeter (1939). In the reality, the two schemes are very different. They do not have much in common. By contrast, the scheme proposed in this section can lead to a unified view of the three cycles. This unified view can be analysed in figure 2.9 where

it is possible to see how the peak in the acceleration cycle anticipates the one of the growth cycle, which again leads the peak of the business cycle. Concerning the troughs, the acceleration cycle is happening first, anticipating the business cycle and then the growth cycle.

Figure 2.9: Schematic Patterns of Growth Cycle and Business Cycle Movements
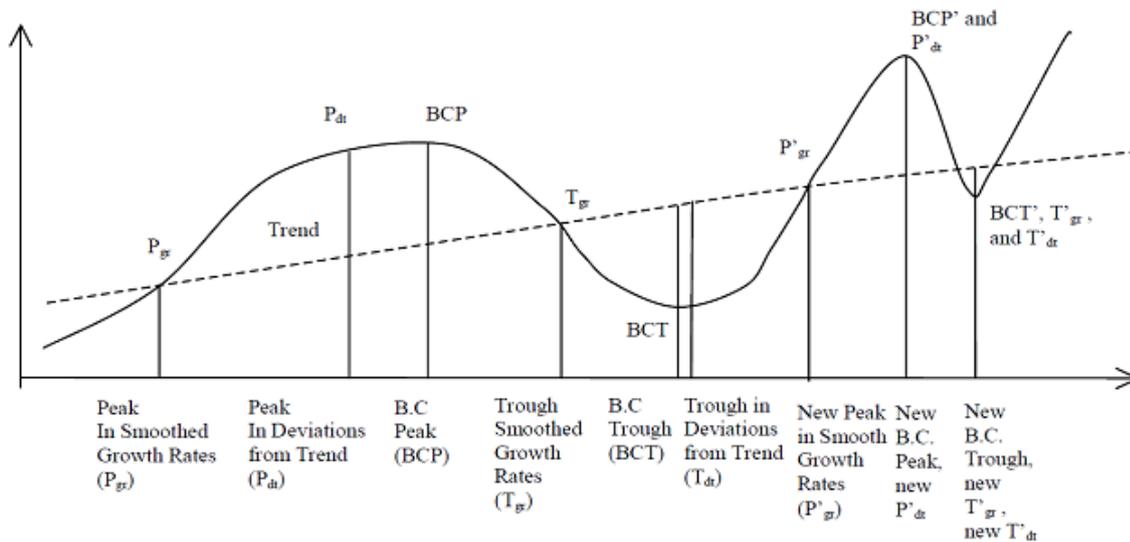
In a complete business cycle fluctuation, growth cycle and acceleration cycle appear to be phases of this fluctuation. Nevertheless, the economy can also decelerate and the re-accelerate without any further consequences on other cycles or it can experience deceleration and slowdown phases formed by acceleration and recovery without any effect in the business cycle. If we define:

- α and β respectively the peak and the trough of the acceleration cycle,

- A and D respectively the peak and the trough of the growth cycle, and

- B and C, and A and D respectively the peak and the trough of the business cycle

then we have the α, A, B, β, C, D scheme proposed by Anas and Ferrara (2004), Anas et al. (2007) and discussed in detail by Anas, Billio, Ferrara, Mazzi (chapter 14).

2.9 Contemporary theories

In this section we aim to provide an overview of main contemporary theories of economic fluctuation organised into two main groups: (i) those which can be seen as an evolution of the Keynesian theory — denominated as neo- and post-Keynesian theories — and (ii) the so-called neo-classical therories (which mainly stem from Friedman and, more generally, the University of Chicago School of Economics).

There is a vast literature related to business cycle models and theories, and we do not pretend here to be exhaustive, but we hope to provide a good picture of the main methodological contributions since the end of the 1940s.

2.9.1 Neo- and post-Keynesian theories

The inventory theory of business cycle: Abramovitz and Stanback

This theory was introduced by Abramovitz (1950) and Stanback (1962) for a simple explanation of short business cycles caused by traders' self-defeating efforts to build or reduce inventories.

The basic assumption of this theory implies that entrepreneurs have a fixed notion of their desired inventory/sales ratio. If demand goes up, the ratio is low and entrepreneurs try to increase it. This effort increases sales and the ratio falls. The upper turning point is the result of the marginal propensity to consume, which is positive, but less than one. Therefore, increases in sales will be smaller than the increases in income. The entrepreneurs try to reduce their inventory levels and their sales fall, and the effort further reduces income and so prevents the fall in the ratio. The rate of decline in sales is lower than the rate of decline in income, and so over time, entrepreneurs re-establish the desired inventory/sales ratio and the contraction bottoms out.

Credit/debt cycle based theories

An alternative theory is the one which assumes that the primary cause of economic cycles is due to the credit cycle: the net expansion of credit (increase in private credit, equivalently debt, as a percentage of GDP) yields economic expansions, while the net contraction causes recessions, and if it persists, depressions. In particular, the bursting of speculative bubbles is seen as the main cause of macro-economic fluctuations. Based on this theory, financial institutions and banks are the main drivers of cyclical movements. A first attempt to formulate a theory in this field is the debt deflation theory of Fisher (1933), which aimed at explaining the Great Depression of 1929.

The most relevant contemporary contribution in this field is the Financial Instability Hypothesis proposed by Minsky (1975, 1982, 1986). The Minsky contribution to explaining economic fluctuations has been very relevant even if it was not much known and recognised, at least during his life. His theory of instability has both micro and macro foundations and making a short synthesis is quite complex. Following Skott (2011) we are providing below a short presentation of the main features of his macroeconomic ideas:

Let us suppose that, after having recovered from a past turbulence period, the economy now appears to be approaching a smooth equilibrium path. Along this path expectations are largely being met and, using Minsky's terminology, there is 'financial tranquillity': borrowers are able to meet their financial commitments. This state of tranquillity will induce changes in the risk assessments of both lenders and borrowers. At the same time, financial regulators and policy makers may loosen regulatory standards. Risk premiums fall; lenders start giving loans they would previously have rejected, and borrowers increasingly finance their projects in speculative and risky ways.

These changes in the behaviour of financial agents relax the financial constraints on the rate of investment and a boom ensues. Gradually, the 'fragility' of the financial system increases until a financial crisis causes a rapid rise in interest rates and a contraction of credit and investment. A return to cautious financial practices now follows and the process repeats itself, although the precise financial instruments and institutions may be new and different. After the 2007-2009 global financial and economic crisis, Minsky's theories, have received a lot of attention and a number of formalizations have been developed such as Ryoo (2010) whose model produces short cycles around a Minskian long wave. Mehrling (1999) has provided a complete bibliography of Minsky papers and an interesting review of his study is proposed by Flanders (2015). Finally, Toporowski (2012) is investigating the link between Kalecki and Minsky monetary theories.

Political based theory of business cycle

Neo-classical macro-economists claim that political intervention is often the cause of cyclical fluctuations so that a neutral attitude is preferable. Within the post-Keynesian theories, one suggests that the behaviour of politicians, either to reinforce their power or to influence elections, can generate cyclical fluctuations. The origin of this theory can be found in a short paper from Kalecki (1943).

Technically speaking, Kalecki (1943) argues, governments may have the ability to control aggregate demand at a level close to full employment. The maintenance of a quasi-full employment condition generates cumulative changes in worker militancy. Increased militancy associated to inflationary pressures generated by the power of workers and the quasi-full employment conditions quickly bring together a powerful block of business entrepreneurs and rentiers. Then the government, also supported by economists who "declare that the situation is manifestly unsound" allows unemployment to rise. The result, Kalecki (1943) argues, is a political business cycle. Although applied by Kalecki (1943) to short cycles, the argument is arguably better suited, as emphasised by Skott (2011), to deal with longer-term fluctuations, and it has been used by a number of writers in relation to the rise in unemployment in the 1970s and 1980s.

In recent years, proponents of the so called "electoral business cycle" theory have argued that incumbent politicians encourage prosperity before elections in order to ensure re-election — and make the citizens pay for it with recessions afterwards. In this field Nordhaus (1975) proposed a model of the so called "opportunistic political cycle" where the party in power stimulates the economy before elections to improve its re-election probability.

Alternatively, the so called "partisan business cycle" suggests that cycles result from the successive elections of administrations with different policy regimes. Regime A adopts expansionary policies, resulting in growth and inflation, but is voted out of office when inflation becomes unacceptably high. The replacement, Regime B, adopts contractionary policies reducing inflation and growth, and the downwards swing of the cycle. It is voted out of office when unemployment is too high, being replaced by Party A. In this field, Hibbs (1977) firstly introduced a partisan business cycle model where a left oriented party was supposed to have one of the following: a higher output target, a higher inflation target, or a higher relative weight on minimizing output rather than a right oriented party.

Several studies have tried to test empirically various political business cycle hypotheses with not very encouraging results. Here we would like to mention, among others: Grier (2008), Alesina et al. (1997), Faust and Irons (1999) . Finally, a comprehensive theoretical and empirical overview of political business cycle theory with an interesting empirical application is provided in Milani (2009).

New Keynesian and Marxist theory of business cycle

In the Keynesian tradition, Goodwin (1951, 1967) developed a growth cycle model of the economy based on the Marxian general law of accumulation. His model accounts for cycles in output by the dynamic interaction between the distributions of income between business profits and workers' wages and the capital accumulation. The fluctuations in wages are almost the same as in the level of employment (wage cycle lags one period behind the employment cycle). When the economy is experiencing a phase of high employment âĂŞ– when the reserve army of labour is small, in Marx's terminology — workers are in a strong position being able to demand higher real wages. In such a situation, profit rates start to suffer and the capital accumulation declines. Assuming a fixed output-capital ratio, the growth rates of output and employment consequently start to decline. In this new scenario the unemployment will start soon to increase implying that the balance of power between workers and capitalist starts shifting against the former. When the workers have lost enough power, the share of wages stop increasing. At this stage, since the level of profitability is low, the rate of accumulation will also be low and the rate of unemployment keeps rising.

The capitalists now can lead the game, the wage share starts falling, and both profitability and accumulation increase. This new situation gradually increases the rate of employment, workers gain again power and

wages start to increase. The Goodwin (1967) cycle is now complete. As stated by Skott (2011) in the original version of the Goodwin (1967) model there are no Keynesian elements since the capital is fully used at any time, the output is determined by the supply side without any role for aggregate demand, and investments adjust in passive way to the level of saving. Several ways to combine Keynesian and purely Marxian aspects into a hybrid model have been proposed following the Goodwin contribution (see Skott (2011)). The Goodwin (1967) model has also been extensively analysed together with some of its possible extension in Weber (2005). Following Weber (2005), the beauty of the models lays on its simplicity being based on the classical Lotka-Volterra predator-prey model.

Despite its attractive features, there are not many empirical studies trying to test the plausibility of the Goodwin model with real data. Among them we notice: Solow (1990), who tested the model on US data, Harvie (2000), who tested the model on 10 OECD countries, and Tarassow (2010).

2.9.2 Neoclassical theories

The new classical theories are attempting to revive classical economics in a way consistent with observed business cycles and yet allowing for market clearing equilibrium. The new Keynesians focus on non-market-clearing models, in which exchange occurs without market clearing prices having been established.

Monetary business cycle theory: Milton Friedman

The Friedman fooling model (Friedman and Schwartz , 1963), based on an imperfect information system of economic agents, serves several purposes:

- it offers the first alternative to the Keynesian assumptions of nominal wage rigidity and non- market-clearing to explain the existence of business cycles, and

- it contains some of the essential elements — market clearing and imperfect information — incorporated into the new classical theory. The innovative feature of Friedman's model is the specification of the labour supply curve to be dependent on the expected real wage rather than the actual real wage. This implies that the presence of imperfect price information on the part of workers will allow the economy to deviate from the long-run natural level of output and generate business cycles.

In fact, in the Friedman and Schwartz (1963) model, expectations are purely adaptive meaning that expectations for the next period's values are based on an average of actual values observed in the previous periods.

Because the level of output is always equal to the natural real GDP when expectations are accurate, the long-run aggregate supply curve is vertical. The model, which obeys the natural rate hypothesis (shifts in aggregate demand have no long-run effect on real GDP) is sometimes called a "natural rate" model. The principal criticism against the Friedman and Schwartz (1963) model is that, since it is very hard to justify that workers can be "fooled" for any great length of time, it doesn't provide a satisfactory theory of business cycles. As well as its importance to the development of modern business cycle theories, the Friedman model helps in the understanding of the new classical model, and of the major issues separating it and the new Keynesian model.

Robert Lucas' theory of business cycle

The model proposed by Lucas (1975, 1977) adds one important element to Friedman's assumption of market-clearing and imperfect information: the assumption of rational expectations (which need not be correct, but make the best use of available information, avoiding errors that could have been foreseen by knowledge of history). People make their best forecasts of the future based on all data currently available rather than having to learn and catch up to the current situation *ex-post*.

Rational expectations (RE) are clearly distinguished from adaptive expectations such as used in Friedman's model. In RE models, individuals are forward-looking and adjust their expectations to their best forecasts of the future. With RE, errors in expectations occur only randomly and independently. RE are incorporated into the Lucas-Friedman supply curve (Gordon , 1976). The reasoning underlying the Lucas supply curve is that whenever the expected price level equals the actual one, the supply curve will be a vertical line at its naturale rate, and that output will only be responsive to changes in prices if the actual price level deviates from the expected level.

The slope of the supply curve can be understood as making a distinction between local and aggregate supply shocks: individual producers will only be willing to supply more if the price of their product rises relative to the general price level (P). Each individual producer is assumed to know the price of its own product, but, because of information barriers, they cannot directly observe the price of other products, thus, for any given price change, they must infer whether this is a local or an aggregate price shock. To the extent that their guess is incorrect, the economy will be able to deviate from the natural level of GDP and generate cyclical fluctuations.

The real business cycle (RBC) theory

There is no doubt that the real business cycle (RBC) theory has constituted the most important macro-economic revolution of the last 40 years. This view is probably shared by both neoclassical economists which see it as a real improvement, but also by Keynesian and post-Keynesian economists who have spent a lot of time and resources in debating the main assumption of the real business cycle theory also by proposing alternative explanations. Consequently, the real business cycle theory (RBC) has been the main protagonist of the economic debate for decades and is still attracting the interest of many researchers.

Giving a concise description of the main points characterising the RBC theory is particularly challenging also because the RBC has substantially changed over the years, moving from a really inflexible model of the economy progressively towards a more flexible structure, able to accommodate also even non-standard neo-classical assumptions.

The RBC originated in the seminal paper of Kydland and Prescott (1982). In this paper, they introduce both a specific theory of the business cycle and a methodology for testing competing theories of business cycle. The main point of the RBC revolution has been the shift between a model of the Walrasian equilibrium, focusing on point equilibrium, to a new concept focusing on a path equilibrium. This step has been the beginning of the dynamic stochastic general equilibrium models (DSGE) which has been largely used and developed in recent years, going even beyond the RBC literature.

The two main assumptions of the Kydland and Prescott (1982) model can be synthesised as follows:

1. Money is of little importance in business cycles.

2. Business cycles are created by rational agents responding optimally to real (not nominal) shocks - mostly fluctuations in productivity growth, but also fluctuations in government purchases, import prices, or preferences. The original model was characterised by an economy evolving on an equilibrium path with rational agents who were maximising their utility. Only a single good was produced according to a given production function and money was not present.

In their paper, Kydland and Prescott (1982) also propose a methodology for evaluating competing business cycle theories. This methodology is based on two main principles:

1. The economy should always be modelled using dynamic general equilibrium models (with rational expectations).

2. The quantitative implications of a proposed model should be taken seriously. In particular, a model's suitability for describing reality should be evaluated using a quantitative technique known as 'calibration'. If the model 'fits' the data, its quantitative policy implications should be taken seriously.

The original version of the RBC theory was strongly criticised by several Keynesian economists such as Summers (1986), Krugman (1994) and Mankiw (1989). Mankiw was questioning all main hypotheses of the real business cycle theory in a quite detailed manner by testing them both theoretically and empirically. An interesting analysis of the RBC models with possible line of extensions has been provided by Stadler (1994).

In empirical studies, the RBC models has demonstrated to be able, when properly calibrated, by using as input technology shocks derived by the residuals of the Solow (1957) model; to reproduce in a good way statistical properties of most economic series. However, they encountered problems in reproducing the co-movements and the cross-correlation structure across macroeconomic variables as demonstrated by Rotemberg and Woodford (1996), King and Rebelo (2000) and Ireland (2003).

RBC models have also proven to be able to accurately reproduce historical cyclical behaviour, provided that an historical sequence of technological shocks has been identified. Unfortunately, this does not happen for all major crises such as the 1982 one or the 2007-2009 for which significant shocks due to technological changes or productivity shocks were not identified. Furthermore, the fact that in the RBC models the cyclical evolution of main macroeconomic variables rely almost exclusively on exogenous productivity shocks, makes them unsuited to economic forecasting. This constitutes a relevant drawback as emphasised by Rotemberg and Woodford (1996).

In the recent years, a number of modified RBC models, also incorporating in some cases certain new-Keynesian hypothesis or some monetary structures, have been proposed in the literature. Their aim was to overcome some rigidities of the original RBC model and to achieve a better description of some business cycle stylised facts.

Following Hallegatte and Ghil (2007), it is possible to identify four main lines on which researchers have introduced changes in the original RBC models:

  (i) varying capital utilization to reproduce realistic cycle amplitudes with small, non-negative changes in productivity (King and Rebelo, 2000);

 (ii) introducing capital constraints to explain cycle asymmetry (e.g., Hansen and Prescott, 2005);

(iii) relying on monopolistic competition, price stickiness and monetary policies to explain business cycle persistence and correlation between nominal and real variables, as presented in e.g. Hairault and Portier (1993); Ellison and Scott (2000); Ireland (2003); Christiano et al. (2005);

(iv) matching friction in the labour market along with wage stickiness to explain the large response of employment to small changes in productivity and the co-movement of output and wages (Christiano et al. (2005); Hall (2005)).

By softening some of the original assumptions and introducing some frictions and market imperfections, new RBC models have lost some of their purely neo-classical attributes, re-joining, at least partially, the new-Keynesian theories. A first consequence is that in these new RBC models, even if business cycle movements are still mainly generated by exogenous factors, also endogenous fluctuations are possible.

To make this evolution clearer, we can — following De Vroey and Pensieroso (2006) — consider how some RBC economists, such as Prekott and Sargent, changed their view on it over the time. Originally, they considered that RBC should not invest in explaining the great depression of the 1930s because it was a unique and not replicable phenomenon. Sargent and Sims (1977) considered that RBC as formulated was unable to explain all elements characterizing the great depression. Finally, Kehoe and Prescott (2002) consider that the great depression was not a unique event but a special situation which occurred not only in the USA in the 1930s but also in most capitalist economies. They also provide a definition of the great depression based on two elements: the size of the deviation from the trend which should be large enough and the rapidity of the deviation. They consider that the great depression should be explained within growth theory abandoning the dogma that it constitutes an aberrant phenomenon not explicable using standard general equilibrium models. This shows how progressively the RBC was moving away from the standard Walrasian equilibrium character-

ising the original RBC models to explore new scenarios incorporating also other non-standard neo-classical features.

Following Romer (2001), it makes sense to distinguish RBC models and RBC style models, where the latter cover a large variety of models deviating in a more or less relevant way from the original specifications.

2.10 The evolution of business cycle tools

As shown in section 2.9, the evolution of tools before the Second World War has been quite moderate, the scene being dominated by the so-called economic barometers. Since they were purely empirical exercises, there was not a clear link between economic theories of cyclical fluctuations and tools for measuring and monitoring them.

Since the end of the 1940s, the situation changed considerably. Since then, we have observed a period of continuous development of new tools for business cycle analysis and monitoring which has never stopped. The reasons for such great changes are due to the large development in econometric theory, time-series analysis and computational methods which started in the second half of the 20th century. Furthermore, new tools started to reflect also the evolution of the business cycle theories or methodologies. An interesting overview of the evolution of econometric tools for business cycle analysis is provided by Qin (2010).

At least in a first phase ending at the beginning of the 1970s, it is possible to identify two main lines of development for business cycle tools. The first line mainly followed the classical econometric approach being based essentially on large scale structural models. This research line was reflecting more or less the Keynesian and neo-Keynesian theories. The second line of research was mainly based on time-series methods and it originated from the works carried out at NBER as well as at some universities. The second line of research has progressively shown some links with the neo-classical theories.

Before going further, it can be of interest mentioning that the large part of the empirical literature on business cycles has originated in the USA, at least until the oil crisis of 1973-1974. This was probably due to the fact that Europe did not suffer particular severe crises, neither before nor after the Second World War. Nevertheless, a lack of data has also played a role in limiting such studies in Europe.

The econometric model of the US economy developed by Tinbergen (1938-39) was the first example of the use of structural econometric models to study and analyse business cycles. Following this model, there was a proliferation of models, especially in the 1950s and 1960s. A methodological summary of econometric modelling of business cycle is provided by Koopmans (1949). These models were generally based on the kind of Frisch (1933) propagation mechanism together with some endogenous generator of cycles. The dynamic structure of these models was relatively simple and their stochastic characteristics quite basic. The use of structural econometric models for business cycle analysis was mainly carried out by means of dynamic simulation exercises based on the work of Adelman and Adelman (1959). A comparative analysis of the most known macroeconometric models has been presented by Hickman (1972). The debate about the use of structural econometric models in business cycle analysis has characterised the 1950s and 1960s until the oil crisis of 1973-1974, which could be considered symbolically as the date of the abandonment of such approaches.

During the same period, the NBER was investing a lot in consolidating its business cycle methodology as developed by Burns and Mitchell (1946). Major attention was put in the improvement of the composite indicators methodology, the study of cycles' characteristics and their relationship across sectors and countries. In this activity, the NBER was studying around 1207 time series, not only from the US economy but also for United Kingdom, France and Germany. The NBER approach has been subject to several critiques, especially from Koopmans (1947), mainly for being based on 'measurement without theory'. Even if the criticisms have some rational foundations, there is no doubt that the NBER research constitutes the core of the contemporary business cycle analysis.

The two main results of NBER research program from the 1950s and 1960s can be considered the formalisation of the first de-trending method by Mintz (1972, 1969) commonly referred as phase average trend (PAT) and the first dating algorithm developed by Bry and Boschan (1971). These developments generated a large variety of contributions in the field of de-trending and turning points dating. These lines of research are extensively discussed in Mazzi et al. (chapter 9) on de-trending methods and in Mazzi (chapter 7 and chapter 13) and Mazzi et al. (chapter 15) on the dating algorithms. For this reason, in this chapter we do not report further on these topics.

A NBER related research was carried out since the end of 1950s at the Princeton University by Morgenstern (1959, 1961). His idea was to study the cyclical propagation across countries (US, UK and France) and to study the link between financial and real cycles (see Morgenstern (1959)). In his study, Morgenstern (1959) was using the NBER definitions and main methodology. During his study he noticed that further improvement was needed in order to perform an accurate investigation. Such improvements should concern both the theoretical aspects as well as the empirical tools used in the comparative studies. Concerning this second aspect he proposed to use spectral and cross-spectral methods in his investigation about the cyclical propagation (see Morgenstern (1961)). This proposal was absolutely innovative for business cycle analysis but it was seen with a certain scepticism by most researchers. In particular Wold (1967) was criticising this approach considering it as "empirism without theory". This criticism was soon neutralized thanks to the research of Granger (1969) on causality and feedback relationships between time series identified by using spectral methods. The causality test and the related assumptions were opening a new era in time series analysis. Granger (1969) who was part of the Morgenstern team made during the 1960s very important contributions to the cyclical analysis in the frequency domain (see also Granger and Hatanaka (1964)).

Since the second half of 1970s, after the failure of structural econometric models to predict the oil crisis, we have observed a growing interest on parametric time-series models. This was determined by several independent factors: the Box and Jenkins (1976) revolution, the Granger causality test and the need for models with a richer dynamic structure and more complex stochastic structure. These three factors have contributed substantially to the development of the vector autoregressive models (VAR) (see Sargent and Sims (1977)). They first explored a dynamic factor model to replicate the latent NBER cycle with unsatisfactory results. They used then a VAR specification with five variables to replicate NBER cycles for the US. In the specification of the model, they were extensively using the Granger causality test in order to identify relationships across variables as well as leading or lagging behaviours.

Generally speaking, VAR models have been appealing research tools for at least three reasons:

1. First, they offer a convenient way to characterise data without having to invoke economic theory to restrict the dynamic relations among variables.

2. Second, many completely specified economic models give rise to VAR representations of the variables in the model. They have then been widely used for both data description and model characterisation.

3. Third, VARs can be readily transformed to interpret the evolution of the system's variables as a function of orthogonalised 'innovations' in any of these variables. Cooley and LeRoy (1985), provide a description of the relation between identification and notions of causality and exogeneity as they apply to VARs.

VARs have been subject to several critricisms mainly because they do not rely on any prior theoretical formulation. However, this has been one of the main advantages of VAR models and probably one of the main reasons for their persisting success. Specifically, concerning their utilisation in the field of business cycle analysis, the drawback can be that the impulse responses they generate do not have any structural interpretation because their innovations are not identified with the underlying structural errors. A response to this problem has been the development of structural vector autoregressive models (SVAR) which have proliferated since the end of 1980s. The development of SVAR models can be attributed to three main factors. The first one being the criticism on VAR models due to their lack of theoretical foundations; the second being identified in the findings of the Blanchard and Watson (1986) work, where they attributed the cyclical fluctuations to a mixture

of large and small shocks, generated in various sectors such as the monetary, the fiscal, the demand and the supply ones. The third, and more technical one, being represented by the generalisation of the Beveridge and Nelson (1981) decomposition to the multivariate case involving a vector of time-series. The Beveridge and Nelson (1981) decomposition was the first de-trending approach moving away from the traditional filtering techniques and exploring potentials of the ARIMA models in this field.

The Beveridge and Nelson (1981) decomposition produced a random walk trend with drift and very irregular cyclical components that accommodates very well the neo-classical hypothesis. Using the Beveridge and Nelson (1981) terminology, it was subsequently referred to as permanent and transitory components, following Friedman (1957). In the SVAR models, firstly proposed by Blanchard and Quah (1989), the problem of interpreting VAR models and identifying properly the impulse response outcomes, is solved by introducing in the model restrictions that are sufficient to identify underlying shocks. The identification is achieved in two stages: in the first one, theoretical restrictions are imposed; they mainly concern the interactions of structural innovations. In the second stage, additional theoretical restrictions are imposed. They reflect the theory of business cycle retained in the model. This set of restrictions allow for a decomposition between permanent and transitory shocks which, respectively, generate the trend and the cyclical components. SVAR models have a Beveridge-Nelson limiting decomposition, while in the short run they tend to behave in a Keynesian-style way. A review of structural VAR models is also provided in Mazzi et al (2016).

Structural VAR models have been subject to many criticisms, especially by neo-classical researchers, mainly because of the difficulty or even the impossibility of testing statistically the imposed restrictions. An overview is provided by Cooley and LeRoy (1985).

Despite the criticism, SVAR models had quite a big success and they have represented an important empirical tool for analysing the effects of a variety of shocks which were thought to generate cyclical movements. In this way, SVAR models have represented a sort of first step in the reconciliation between Keynesian and neo-classical views, at least because they were offering a common modelling scheme where a variety of shocks could be hosted and analysed. Furthermore, the relations between DSGE and SVAR models, as analysed by Christiano et al. (2007), have contributed to enhance the relevance of SVAR models.

The 1980s represented also a period where VAR models assumed increasing importance in relation to newly discovered characteristics of economic data, such as the presence of co-integration across variables, etc. New forms of VAR models, such as the vector error correction model (VECM) started to become popular in the context of business cycle analysis. At the same time, dynamic factor models started to be used to replicate the latent reference cycle of NBER, for example by Stock and Watson (1989). Finally, the use of DSGE models, initiated by Kydland and Prescott (1982), started to become a new dominant modelling tool, revitalising in a way the structural macroeconomic modelling. As already mentioned before (Romer , 2001, see also), DSGE models started soon to gain a certain degree of independence from the real business cycle theory, accommodating also post-Keynesian hypotheses. In this way, they completed the reconciliation between Keynesian and neo-classical views, already started with structural VAR models, by offering a common modelling framework to analyse and compare several ways of generating economic cycles. The tools mentioned in the last part of this section still play a crucial role in the contemporaneous business cycle analysis and they are considered, not only as competing, but also as complementing each other in order to give a better description of business cycle movements.

2.11 New challenges after the global economic and financial crisis

The popular statement that each crisis is different from the other has once again proven its truthfulness with the 2008-2009 global financial and economic crises as well as with the 2012-2013 sovereign debt crisis in

Europe. Each big crisis has generated significant changes and evolutions in the theories of business cycle as well as in the tools used for monitoring and analysing them.

At this stage, it is probably too early to identify the repercussions of the two crises of the 21st century both for the theoretical and empirical aspects. Nevertheless, it is already possible to identify the main elements which appear to differentiate them from the previous crises. Such differentiating elements can be synthesised as follows:

- the central and crucial role played by globalisation which has probably transformed the business cycle analysis from a country specific to a global specific issue;

- the strong link between financial and real cycles observed in the last crisis;

- the role of speculation and bubbles in specific sectors such as the housing market;

- the interaction between structural macroeconomic imbalances, structural reforms, and cyclical movements.

Business cycle theories and the associated statistical and econometric tools need to incorporate such new elements in order to be ready to the challenges ahead.

2.12 Conclusions

This chapter has provided an historical overview of the evolution of business cycle theories and of the associated empirical tools since the end of the 19th century until the end of the 20th century. We have followed, as much as possible, in chronological sequence by alternating the presentation of theories with the presentation of tools available contemporaneously. Obviously, this sequence cannot be perfect because theories and tools did not evolve perfectly in parallel, especially until the Second World War. Since the topic of the chapter is very large, it was impossible to provide a complete exhaustive picture of theories and tools but we have tried to present the most relevant ones by providing detailed references which could be beneficial for the readers willing to go more in detail.

Bibliography

Abramovitz, M. (1950), Inventories and business cycles, with special reference to manufacturers' inventories. New York: NBER.

Adams B., Bottelier P., Ozyildirim A., Sima-Friedman J. May 2010 - EPWP 10 - 02.

Adelman, F. and Adelman, I. (1959), The dynamic properties of the Klein-Goldberger Model, Econometrica, 27, 596-625.

Allgoewer E. (2002), Underconsumption theories and Keynesian economics, Interpretations of the Great Depression, May 2002, Discussion paper no. 2002-14.

Åkerman J. (1928), Om det ekonomiska livets rytmik, Diss. Lund : University of Stockholm.

Åkerman, J. (1932), Economic progress and economic crisis, Macmillan, London.

Alesina, A.,N.Roubini, and G.Cohen (1997), Political Cycles and the Macroeconomy, Cambridge, MA:MITPress.

Anas J., Ferrara L. (2004), Turning points detection: The ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, No. 1, Vol. 2, pp. 1-36.

Anas J., Billio M., Ferrara L., Mazzi G.L. (2007), A System For Dating And Detecting Turning Points In Tthe Euro Area, Manchester School, University of Manchester, vol. 76(5), pages 549-577, 09.

Armatte M. (1992), Conjonctions, conjoncture, et conjecture. Les baromètres économiques (1885-1930), Histoire & Mesure, 1992 volume 7 - n° 1-2. Histoire de la pensée économique. pp. 99-149.

Armatte M. (2003), Indicateurs économiques - L'indice des prix: un objet de controverses - La Recherche, Problèmes économiques : le meilleur de la presse et des revues pour suivre l'actualité (2003).

Artis M.J., Kontolemis Z.G., Osborn D. (1995), Classical Business Cycles for G7 and European Countries, CEPR Discussion Papers 1137, C.E.P.R. Discussion Papers.

Beaudry P., Portier F. (2004), When Can Changes in Expectations Cause Business Cycle Fluctuations in Neo-Classical Settings? CEPR Discussion Papers 4628, C.E.P.R. Discussion Papers.

Beveridge W.H. (1909), Unemployment, a problem of industry, 1re ed. 1909, 2e édition 1931, Londres.

Beveridge S., and Nelson C. R. (1981), A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle." J. Monet. Econ. 7:151-74.

Blanchard O. J., and Quah D. (1989), The dynamic effects of aggregate demand and supply disturbances. Amer. Econ. Rev. 79(4):655-73.

Blanchard O. J., and Watson M. W. (1986), Are business cycles all alike?, In Gordon 1986a.

Boschan, C., Ebanks, W. W. (1978), The phase-average trend: A new way of measuring growth. In 1978 proceedings of the Business and Economic Statistics Section. Washington, D.C.: American Statistical Association.

Box, G. E. P., and Jenkins, G. M. (1976), Time series analysis, forecasting, and control, San Francisco: Holden-Day.

Bry G., and Boschan C. (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, New York, NBER.

Burns A. F., and Mitchell W. C. (1946), Measuring business cycles. New York: NBER.

Carry A. (2003), Kondratieff or how to convince people that long-term economic exists, Monographs of official statistics, Papers and proceedings of the colloquium on the history of business-cycle analysis, Luxembourg: 2003.

Christiano L., Fitzgerald T.J. (2003), The Band Pass Filter, International Economic Review, 44, 2, 435-465.

Christiano, L.J., M. Eichenbaum, and C.L. Evans, (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, Journal of Political Economy 113, 1-45.

Christiano, L.J., M. Eichenbaum, and R. Vigfusson (2007). Assessing Structural VARs, NBER Macroeconomics Annual 2006, Volume 21.

Collard D.A. (1996), Pigou and modern business cycle theory, The Economic Journal, 106, 912-924.

Cooley, T. F., and LeRoy, S. (1985), Atheoretical macroeconometrics: a critique, Journal of Monetary Economics 16, 283-308.

Cooley, T. F., and Dwyer, M. (1998), Business cycle analysis without much theory: A look at structural VARs, Journal of Econometrics 83 (1998), 57-88.

Cottrell A. (1993), Keynes's Theory of Probability and Its Relevance to His Economics: Three Theses, Economics and Philosophy 9 (01):25 (1993).

Dal-Pont Legrand M., Hagemann H. (2005), Business Cycles in Juglar and Schumpeter, European Society for the History of Economic Thought, University of Stirling, 9-12 June 2005.

Diebold, F. X., Rudebusch G. D. (1989), Scoring the leading indicators, J. Business 62(3):369-91.

Diebold F. X., Rudebusch G. D.(1996), Measuring Business Cycles: A Modern Perspective, Review of Economics and Statistics, vol., LXXVIII, no. 1, February 1996, pp . 67-77.

Domar E. (1946), Capital Expansion, Rate of Growth, and Employment, Econometrica 14 (2): 137–147.

De Foville A. (1888), Essai de météorologie économique et sociale, JSSP, May 1888, p.243.

De Vroey M.R., Pensieroso L. (2006), Real Business Cycle Theory and the Great Depression: The Abandonment of the Abstentionist Viewpoint, Contributions to Macroeconomics, Volume 6, Issue 1, Article 13.

Ellison M., Scott A. (2000), Sticky prices and volatile output, Journal of Monetary Economics 46, 621-632.

Observations on the Barometer and Thermometer, and Account of the Whole Rain in Every Month of the Year 1767, Taken at the Royal Hospital Near Plymouth: Transmitted to William Watson, M. D. F. R. S., Philosophical Transactions (1683-1775), Volume 58, pp. 136-139.

Faust, J., and J. Irons (1999), Money, Politics, and the Post-War Business Cycle, Journal of Monetary Economics, 43, 61-89.

Fayole J. (2003), The study of cycles and economic analysis in the history of economic thought, Monographs of official statistics, Papers and proceedings of the colloquium on the history of business-cycle analysis, Luxembourg: 2003.

Festré A. (2009), Knut Wicksell and Ludwig von Mises on money, interest and price dynamics, Journal of the History of Economic Thought 23, 3 (2006) 333-357.

Filardo A.J. (1994), Business-Cycle Phases and Their Transitional Dynamics, Journal of Business & Economic Statistics Vol. 12, No. 3, Jul., 1994.

Fisher I. (1933), The debt-deflation theory of great depressions, Econometrica 1:337- 57.

Flanders M. J. (2015), It's Not a Minsky Moment, It's a Minsky Era, Or: Inevitable Instability, Econ Journal Watch 12(1).

Forget E.L. (1990), John Stuart Mill's business cycle, History of Political Economy, 22 (4), 629–642.

Foster W.T., Catchings W. (1925), Profits, Boston and New York: Houghton Mifflin Company.

Friedman, M. (1957), A theory of the consumption function. Princeton, N.1.: Princeton University Press for NBER.

Friedman, M., Schwartz. A. (1963), Money and business cycles, Rev. Econ. Statis., 45, suppl.:32-64.

Friedman, M., Schwartz. A. (1963) A monetary history of the United States, 1867-1960, Princeton, N.J.: Princeton University Press for NBER.

Frisch R. (1933). Propagation and Impulse Problems in Dynamic Theory, in Essays in Honor of Gustav Cassel, George Allen and Unwin, London.

Geiger N. (2012), Cycles versus growth in Schumpeter a graphical interpretation of some core theoretical remarks, XIV Colloque de l'Association Charles Gide pour l'étude de la pensée économique.

Goodwin, R. M. (1951), The non-linear accelerator and the persistence of business cycles, Econometrica 19: 1-17.

Goodwin, R. M. (1967), A growth cycle. In Socialism, capitalism, and economic growth: Essays presented to Maurice Dobb, ed. C. H. Feinstein. Cambridge: Cambridge University Press.

Gordon, R. J. 1976. Recent developments in the theory of inflation and unemployment. Journal of Monetary Economics, 2: 185âĂŞ219.

Granger C. W. J., Hatanaka M. (1964), Spectral analysis of economic time series, Princeton, N.J.: Princeton University Press.

Granger C. W. J. (1969), Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424-38.

Grier K. (2008), US presidential elections and real GDP growth, 1961âĂŞ2004, Public Choice, 2008, vol. 135, issue 3, 337-352.

Haberler G. (1946), Prosperity and depression A Theoretical Analysis of Cyclical Movements, United Nations Lake Success, New York 1946.

Hairault, J. and Portier, F. (1993). Money, New-Keynesian macroeconomics and the business cycle, European Economic Review, Elsevier, vol. 37(8), pages 1533-1568.

Hall R. E. (2005), Employment fluctuations with equilibrium wage stickiness, The American Economic Review, 95(1), 50-65.

Hallegatte, S. and Ghil, M. (2007), Endogenous Business Cycles and the Economic Response to Exogenous Shocks, FEEM Working Paper No. 20.2007.

Hamilton J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle, Econometrica 57, 357-384.

Hamilton J.D., Perez-Quiros G. (1996), What do the Leading Indicators Lead?, Journal of Business. Vol 69, Number 1. January 1996.

Hansen G.D., Prescott E.C. (2005), Capacity constraints, asymmetries, and the business cycle, Elsevier for the Society for Economic Dynamics, vol. 8(4), pages 850-865, October.

Harrod R.F. (1939), An Essay in Dynamic Theory, Economic Journal, Vol. XLIX, March 1939.

Harvie, D. (2000). Testing Goodwin: growth cycles in ten OECD countries. Cambridge Journal of Economics, 24, 349-376.

Hawtrey R.G. (1913). Good and Bad Trade. Longmans: London.

Hawtrey R.G. (1919). Currency and Credit. Longmans, Green and Co: London.

Hawtrey R.G. (1928). Trade and Credit. Longmans: London.

Harrod R.F. (1939). An Essay in Dynamic Theory, Economic Journal, Vol. XLIX, March.

Hayek F. A. (1935). Prices and Production, revised edition, London, Routledge.

Hayek F. A. (1966). Monetary Theory and the Trade Cycle, New York, Augustus Kelley.

Heertje A., PHeemeijer P. (2002), On the Origins of Samuelson's Multiplier-Accelerator Model, History of Political Economy, 34 (1): 2007-2018.

Hibbs, D. Political Parties and Macroeconomic Policy, American Political Science Review, 71, 1467-87.

Hickman, B.G. (1972), Econometric Models of Cyclical Behavior, New York, Columbia University Press.

Hicks J. (1939), Value and Capital, Oxford: Clarendon.

Hobson J.A. (1933). Under-consumption. An exposition and a reply., Economica, VOL XIII (old series), 402–27.

Hodrick, R. J., and E. C. Prescott (1980), Post-war U.S. business cycles: An empirical investigation, Discussion Paper no. 451, Carnegie-Mellon University, Pittsburgh.

Ireland P.N. (2003), Endogenous money of sticky prices? Journal of Monetary Economics 50, 1623-1648

Jevons W.S. (1866), Brief Account of a General Mathematical Theory of Political Economy, The Journal of the Royal Statistical Society, London, XXIX (June 1866), pp. 282-87. Section F of the British Association, 1862.

Jevons W.S. (1875), Money and the Mechanism of Exchange, New York: D. Appleton and Co.

Jevons W.S. (1884), Investigations in currency and finance, Macmillan and Co, London

Juglar C. (1862), Des Crises commerciales et leur retour periodique en France, en Angleterre, et aux Etats-Unis. Paris: Guillaumin, 1862.

Julin A. (1913), La sémiologie statistique, BUS, 1913

Kaldor, N. (1940), A model of the trade cycle, Econ. J. 50:78-92.

Kaldor, N. (1954), The Relation of Economic Growth and Cyclical Fluctuations, The Economic Journal.

Kalecki, M. (1943), Political Aspects of Full Unemployment, Political Quarterly, V. 14 (Oct.-Dec.): 322-331.

Kehoe T. J., and Prescott E. C. (2002), Great Depressions of the 20th Century, Review of Economic Dynamics, vol. 5, pp. 1-18.

Keynes J.M. (1936), The General Theory of Employment, Interest and Money, London: Macmillan (reprinted 2007).

King R. G., and Rebelo S. T. (2000), rescuscitating Real Buisnes Cycles, Working Paper 7534, NBER Working Paper Series.

Klein P. A., Moore G. H. (1985), Monitoring growth cycles in market-oriented countries: Developing and using international economic indicators, Cambridge, Mass.: Ballinger for NBER.

Kitchin J. (1923), Cycles and trends in economic factors, Rev. Econ. Statis. 5: 10-16.

Kondratieff N. D. (1926), Die langen Wellen der Konjunktur, Archiv fur Sozialwissenschaft und Sozialpolitik 56(3):573-609. Also, translated by W. F. Stolper as The long waves in economic life, Rev. Econ. Statis. 17(6)(1935):105-15.

Koopmans, T. C. (1947), Measurement without theory, Review of Economic Studies 29, 161-72.

Koopmans, T. C. (1949), The econometric approach to business fluctuations, American Economic Review 39, 64-72.

Korotayev A.V., Tsirel S.V. (2010), A Spectral Analysis of World GDP Dynamics: Kondratieff Waves, Kuznets Swings

Krugman P. (1994), Peddling Prosperity: Economic Sense and Nonsense in the Age of Diminished Expectations. New York: W.W. Norton & Company.

Kydland, F., and E. C. Prescott (1982). Time to build and aggregate fluctuations, Econometrica 50:1345-70.

Kuznets S. (1940), Schumpeter's Business Cycles, The American Economic Review.

Lucas R.E. Jr. (1975), An equilibrium model of the business cycle. J. Polito Econ. 83:111344.

Lucas R.E. Jr. (1977), Understanding business cycles. In Brunner and Meltzer 1977.

Malthus T.R. (1820), Principles of Political Economy: Considered with a View to Their Practical Application (London: John Murray). Second edition 1836.

Mankiw, N.G. (1989), Real Business Cycles: A New Keynesian Perspective, Journal of Economic Perspectives -Volume 3, Number 3 -Summer 1989 - Pages 79-90.

Marrama, V. (1946-7), Short Notes on a Model of the Trade Cycle, The Review of Economic Studies.

Matthews R.C.O. (1984). Animal spirits, Proceedings of the British Academy, 70, 209-229.

Mazzi G.L., Mitchell J., and Moauro F. (2016), Structural vector autoregressive (SVAR) based estimates of the euro area output gap: theoretical considerations and empirical evidence, Eurostat Statistical Working Papers, Publications Office of the European Union.

Mehrling P.(1999), The Vision of Hyman P. Minsky. Journal of Economic Behavior & Organization 39: 129-158.

Milani, F. (2009), Political Business Cycles in the New Keynesian Model, Economic Inquiry.

Mill J.S. (1826), Paper currency and commercial distress, The collected works of J.S. Mill edited by J.M. Robson (4:71-123, Toronto).

Minsky, H. P. (1975), John Maynard Keynes, New York: Columbia University Press.

Minsky, H. P. (1982), Can "it" happen again? Essays on instability and finance. New York: M. E. Sharpe.

Minsky, H. P. (1986), Stabilizing an Unstable Economy, New Haven: Yale University Press.

Mintz, I. (1969), Dating postwar business cycles: Methods and their application to western Germany, 1950-67, Occasional Paper no. 107. New York: NBER.

Mintz, I. (1972), "Dating American growth cycles." in Zamowitz "The business cycle today". New York: NBER, 1972.

von Mises L. (1928), Monetary Stabilization and Cyclical Policy, German edn., Yena. Reproduced in P. Greaves (Ed)., On the Manipulation of Money and Credit (Free Mkt. Books, 1978).

von Mises L. (1934), The Theory of Money and Credit, translation by H. E. Batson (London: Jonathan Cape Ltd).

von Mises L. (1936), The 'Austrian' Theory of the Trade Cycle, in S. Littlechild (Ed), Austrian Economics, vol. II (Aldershot: Edward Elgar, 1990, pp. 291-96).

Mitchell W. C., Burns A. F. (1938), Statistical indicators of cyclical revivals, Bulletin no. 69, New York: NBER.

Moore H.L. (1914), Economic Cycles: Their Law and Cause, New Pork: The Macmillan Company.

Moore H.L. (1923), Generating Economic Cycles, New Pork: The Macmillan Company.

Morgenstren, O. (1959), International Financial Transactions and Business Cycles, Princeton: Princeton University Press.

Morgenstren, O. (1961), A new look at economic time series analysis, in H. Hegeland ed. Money, Growth, and methodology and other essays in economics: in honor of johan Akerman, CWK Gleerup Publishers, Lund, pp.261-72.

Mourre C. (1913), La prévision des crises commerciales, JSSP, 1913, p. 202–212.

Neftci, S. N. (1992), Are economic time-series asymmetric over the business cycle?, J. Polito Econ. 92(April):305-28.

Neumann-Spallart (1887), La mesure des variations de l'état économique des peuples, read on 13 April 1887, Bulletin de l'I.I.S., II, p.150.

Nordhaus, W. (1975), The Political Business Cycle, Review of Economic Studies, 42, 1975, 169-90.

Orphanides and Van Norden S. (1999). The reliability of output gap estimates in real time, FEDS discussion paper.

Orphanides and Van Norden S. (2002). The unreliability of output gap estimates in real time, The review of Economics and Statistics, 84, 569-83.

Okun A. (1962), Potential GNP, its measurement and significance, Cowles Foundation, Yale University.

Pasinetti L. (1960), Cyclical Fluctuations and Economic Growth, Oxford Economic Papers, New Series, Vol. 12, No. 2, pp. 215-241.

Pigou A. C. (1927), Industrial fluctuations, London: Macmillan.

Persons W. (1916), The construction of a Business Barometer: annual data, AER.

Persons W. (1919), An Index of General Business Conditions, Review. of Economic Statistics, Vol. 1, April, p.111–205.

Qin D. (2010), Econometric Study of Business Cycles in the History of Econometrics, working paper no 669, July 2010.

Robertson, D. H. (1915, 1948), A study of industrial fluctuations, London: Aldwych.

Romer, D. (2001), Advanced Macroeconomics, McGraw-Hill, New York.

Rotemberg, J. J., and Woodford, M. (1996), Real Business Cycles Models and the Forecastable Movements in Output, Hours amd Consumption, American Economics Review 86:71 89.

Ryoo, S. (2010), Long waves and short cycles in a model of endogenous financial fragility. Journal of Economic Behavior and Organization, 74, 163-186.

Samuelson P. A. (1939a), Interactions between the Multiplier Analysis and the Principle of Acceleration, Review of Economics and Statistics 21:75–78.

Samuelson P. A. (1939b), A Synthesis of the Principle of Acceleration and the Multiplier, Journal of Political Economy 47:786–97.

Sargent T. J., Sims C. A. (1977), Business cycle modeling without pretending to have too much a priori economic theory, New methods in business cycle research: Proceedings from a conference, Minneapolis: Federal Reserve Bank of Minneapolis.

Schumpeter J.A. (1939), Business Cycles, A theoretical, Historical, and Statistical Analysis of the Capitalist Process, New York, McGraw-Hill.

de Sismondi J.C. (1819), Nouveaux Principes d'économie politique.

Skott, P. (2011), Post-Keynesian theories of business cycles, Working Paper 2011âĂŘ21, University of Massachusetts Amherst.

Solow, R.M. (1957). Technical change and the aggregate production function, Rev. Econ. Statis. 39(August):312-20.

Solow, R.M. (1990). Goodwin's Growth Cycle: Reminiscence and Rumination. In K. Velupillai, ed. Nonlinear and Multisectoral Macrodynamics: Essays in Honour of Richard Goodwin. New York: University Press, pp. 31-41.

Stadler, G.W. (1994). Real Business Cycles, Journal of Economics Literatute, Vol. XXXII (December 1994), pp. 1750âĂŞ1783.

Stanback, T.M.Jr. (1962), Postwar cycles in manufacturers' inventories. New York: NBER.

Stock, J. H., and M. W. Watson (1989), New indexes of coincident and leading economic indicators, NBER Macroeconomic Annual, 351-409.

Stock, J. H., and M. W. Watson (1991), Predicting recessions., Paper presented at the NBER conference on New Research on Business Cycles, Indicators, and Forecasting, May, Cambridge, Mass.

Stock, J. H., and M. W. Watson (1991), A probability model of the coincident economic indicators. In Lahiri and Moore 1991.

Summers, L.H. (1986), Some Skeptical Observations on Real Business Cycle Theory, Federal Reserve Bank of Minneapolis, Quarterly Review, Fall 1986, 23-27.

Tarassow, A. (2010), The empirical relevance of Goodwin's business cycle model for the US economy, Profitanalysegruppe (PRAG) am Zentrum für Ökonomische und Soziologische Studien (ZOESS), Discussion Papaer No 5.

Tinbergen, J. (1938-39), Statistical testing of business-cycle theories. 2 vols. Geneva: League of Nations, Economic Intelligence Service.

Toporowski, J. (2012), The Monetary Theory of Kalecki and Minsky, SOAS Department of Economics Working Paper Series, No. 172, The School of Oriental and African Studies.

Thorp W.L. (1926), Business Annals, (National Bureau of Economic Research, New York).

Weber L. (2005), A contribution to Goodwinâ ĂŹs growth cycle model from a system dynamics perspective, available from (http://www.systemdynamics.org/conferences/2005/proceed/papers/WEBER196.pdf).

Wicksell, K. (1898, 1936), Interest and Prices, Kahn, New York: M. Kelley Publishers.

Wold, H.o. (1967), Book review: C.W.J. Granger (in association with M. Hatanaka) Spectral analysis of economic time series, The Annals of mathematical Statistics, 38, 288-93.

Zarnowitz V., Ozyildirim A. (2000), Time Series Decomposition and Measurement of Business Cycles, Trends and Growth Cycles, Joint Statistical Meetings - Business & Economic Statistics Section.

3

Definitions and Taxonomy of Indicators

Authors

Andrea Carriero
Queen Mary, University of London
a.carriero@qmul.ac.uk

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Massimiliano Marcellino
Bocconi University, IGIER, and CEPR
massimiliano.marcellino@unibocconi.it

Table of Contents

3.1 Introduction

Monitoring the current status of the economy and anticipating its future trends are very relevant for policy making but also for the decisions of private agents, consumers and firms. Unfortunately, there is no consensus in the literature on the selection of a measure of the current and the future status of the economy, i.e., a unique statistical indicator.

By slightly generalising the requirements for leading indicators presented in the work of Moore and Shiskin (1967) to cyclical composite indicators, it is possible to list a set of properties that CCI should exhibit. They include: (i) consistent timing either as a leading, coincident or lagging indicator (i.e., to systematically lead, coincide or lag peaks and troughs in economic activity); (ii) economic significance (i.e., being supported by economic theory as good measures of economic activity); (iii) statistical reliability of data collection (i.e., provide an accurate measure of the quantity of interest); (iv) prompt availability without major later revisions (i.e., being timely and regularly available, without requiring subsequent modifications of the initial statements); (v) smooth month to month changes (i.e. being free of major high frequency movements).

Given these requirements, a natural choice as a coincident indicator is gross domestic product (GDP) or its growth rate, since it is typically considered as a reliable summary of the current economic conditions. Unfortunately, GDP is not available on a monthly basis and, although both in the US and in Europe there is a growing interest in increasing its sampling frequency from quarterly to monthly, the current results are still too preliminary to rely on. Moreover, there are typically long delays in the release of GDP data which is only partially mitigated by the growing use of rapid estimates techniques. Also, preliminary values can be subject to subsequent, even large, revisions which tend to increase with timeliness improvements. Both these features of the GDP data production process make it hardly usable as a timely coincident indicator.

In the past, industrial production (IP) provided a good proxy for the fluctuations of GDP, and it is still currently monitored, for example, by the NBER business cycle dating committee and by the Conference Board (CB) in the US in conjunction with other indicators. Yet, the ever rising share of services compared with the manufacturing, mining, gas and electric utility industries casts more and more doubts on the usefulness of IP as a single coincident indicator. Furthermore, the IP is often characterised by a high degree of volatility which makes the identification of a proper and reliable signal complicated.

Another commonly considered statistical indicator is the volume of sales of the manufacturing, wholesale and retail sectors, deflated for price changes so as to proxy real total spending. Its main drawback, as in the case of IP, is its partial coverage of the economy.

A variable with a close to global coverage is real personal income less transfers, which underlies consumption decisions and aggregate spending. Yet, it is seldom available for European countries. Moreover, unusual productivity growth and favorable terms of trade can make income behave differently from payroll employment, which is the other most common indicator with economy wide coverage. Some authors focused on unemployment rather than employment, e.g. Boldin (1994) or Chin et al. (2000), on the grounds that the series is timely available and subject to minor revisions. Yet, typically unemployment is lagging and persistent rather than coincident, particularly in Europe.

Overall, it is difficult to identify a single variable which provides a good measure of current economic conditions, that is available on a monthly basis, and satisfies all the requirements listed above. Therefore, it is preferred to consider a combination of several statistical indicators, i.e., a cyclical composite indicator (CCI).

In this chapter we survey alternative methodological approaches to CCI construction. As an illustration, we implement some of them for the four largest countries in the euro area, and for the euro area as a whole, complementing the results in Marcellino (2006) for the US and in Carriero and Marcellino (2007) for the UK.

In Section 3.2 we briefly present a taxonomy of cyclical composite indicators; section 3.3 reviews the econometrics of these alternative methods for the construction of a CCI. We also suggest a new pooling based approach. Since constructing a CCI can be considered as a problem of estimation of missing observations

(about the status of the economy), the good performance of pooling emerging from the forecasting literature suggests that combining a set of competing CCIs can improve upon the quality of each of the single CCIs. The analysis is complicated by the fact that the forecasts can be compared with realized values after some time, while the status of the economy remains unobservable. This, somewhat, limits the range of feasible pooling techniques, as in the case of backdating or interpolation, see Marcellino (2007). Yet, simple combination methods such as averaging, possibly after trimming extreme values, work quite well in practice when compared with more sophisticated techniques, see e.g. Stock and Watson (1999) and Marcellino (2004).

In Section 3.4 we apply the alternative techniques for constructing CCIs for France, Germany, Spain and the UK. Since it is difficult to select a methodology from a theoretical point of view, we wish to consider whether they lead to major differences in the evaluation of the status of the economy, or whether a consensus can be achieved, as in the case of the US, see Marcellino (2006), and of the UK over a shorter sample, see Carriero and Marcellino (2007). For simplicity and for the sake of comparability, rather than discussing in details the selection of the components of the CCIs for each country, we rely whenever possible on the variables included into the Conference Board coincident indexes for these European countries. The specific components of the index for each country are also discussed in Section 3.4.

Finally, Section 3.5 concludes with a summary of the main results we have obtained.

3.2 Taxonomy of cyclical composite indicators

CCIs are a large family of indicators measuring different aspects of economic activity. They have different timings, serve various purposes, are based on a variety of statistical/econometric methods, and rely on a large set of data, stemming from quantitative statistical variables to qualitative ones to financial indicators etc.

It might be of interest to present several ways to classify CCIs according to a given number of criteria.

3.2.1 CCI by timing

The most intuitive way of classifying CCIs is, according to their timing, into: leading, coincident and lagging ones. Leading indicators are those anticipating the future pattern of the economy, while coincident ones are those describing the current pattern of economic situation. By contrast, lagging indicators are intended to reproduce, today, the past pattern of the economy. It can be useful to underline as coincident indicators can often fill gaps in official statistics, which can be available usually with a certain time lag so that there are no information for the current period. When compiling composite coincident and leading indicators it is important to pay particular attention on timeliness of official statistics by avoiding the use of variables, which are available too late, and to find the most appropriate mixture between quantitative and soft data (opinion surveys, financial data, etc.) in order to obtain the most accurate description of the current and near future economic trends. The CLI (composite leading indicator) produced by the OECD for its member countries and the leading index elaborated by the Conference Board for the US economy, as well as for other relevant economies around the world, are typical examples of composite leading indicators. The Euro-coin, previously released by the Centre for Economic Policy Research (CEPR) and now, in its new version, released by Bank of Italy, is an example of the GDP monthly coincident indicator for the euro area.

Other examples of coincident indicators, but conceptually completely different, are the purchasing manager index (PMI) elaborated for the US, euro area and some member states, along with the economic sentiment indicator (ESI) compiled by the Directorate General for Economic and Financial Affairs of the European Commission (DG ECFIN) for the euro area, the European Union and its member states. Those two indicators are based on monthly surveys and they represent, in a qualitative way, the opinion of managers (PMI) and of enterprises and households (ESI). The relevance of such indicators is strengthened by their very high timeliness This is particularly true for the ESI as it is available before the end of the reference month, while the PMI is

available a few days after its end. It is more complex to find publicly available examples of lagging indicators since they are usually compiled and used for validation purposes of coincident and leading indicators.

3.2.2 CCI by targeted pattern

CCI can measure different aspects of the economic situation by targeting either the cyclical component or the short-term economic evolution. Also according to the classification used in this handbook, we propose to split the cyclical composite indicators in three main groups: cyclical indicators, turning points indicators (recession indicators) and indicators measuring the short-term economic evolution. Obviously, all those indicators can have leading, coincident or lagging properties. It is important to point out that cyclical and turning point indicators mainly refer to unobserved components, while indicators measuring the short-term economic evolution are referred to observed measures such as the GDP growth or the inflation rate. For instance, the Conference Board leading index is a typical cyclical indicator targeting the business cycle, whereas the OECD CLI is targeting the growth cycle and it is optimised to anticipate its turning points. See Ozyildirim chapter 9 and Gyomai et al., chapter 10, for more details on those two indicators.

At the European level, publicly available turning points indicators (or recession indicators), are IARC and IESR released by ACE (France), (see Anas and Ferrara (2002)). Furthermore, a large number of turning points indicators have been developed during the last decade, also due to the global financial and economic crisis. In this context, we can mention Chauvet and Hamilton (2006) Harding and Pagan (2011). Most of those indicators refer to the business cycle but J. Anas, M. Billio, L. Ferrara, GL. Mazzi, provide turning points indicators for both the business cycle and the growth cycle, either in an univariate or a multivariate framework. Finally, the already mentioned Euro-coin indicator, as well as the publicly available GDP leading indicator for the euro area Euroframe (released by a consortium of European research institutes) and the IRC (produced by ACE, France) are examples of indicators aiming to estimate the current and the future short-term economic evolutions. Furthermore, in this context it is important to mention a contribution of Camba-Méndez and Kapetanios (2004), Camba-Mendez et al. (2011) and G. Camba-Mendez, G. Kapetanios and M. Weale (chapter 16). Furthermore, there was relevant literature on the use of large-scale factor models to nowcast GDP and/or inflation, of which examples are Forni et al. (2001), Banbura et al. (2011), Banbura et al. (2012a) and Banbura et al. (2012b). A detailed analysis of the use of the large-scale factor models is also provided by Luciani (chapter 17). On the other hand, the use of small-scale factor model, in combination with other statistical tools such as linear regression, VAR models or unobserved component models has been explored, with reference mainly to the euro area GDP growth, by Charpin et al. (2008), Charpin et al. (2012), Charpin (chapter 18) and Frale et al. (2010) and Frale et al. (2011).

3.2.3 CCI by kind of compilation method

In this subsection, we focus our attention on classifying cyclical composite indicators according to the compilation method used. When classifying the indicators we can consider only the final step of the procedure related to the method used for the construction of the indicator or also include the variable and model selection process. When restricting to the first case, we can say that, in a very general way, CCIs can be constructed either by using aggregation schemes or model based approaches. In the first category, CCIs are constructed by simple weighted averages of selected statistical indicators. Examples are provided by the Conference Board; see Ozyildirim (chapter 8) and Economic Cycle Research Institute (ECRI) CCIs. Obviously, the selection of compilation methods, and of the weighting scheme, can be made in very different ways. For example, weights can be derived in a purely subjective way or by using some statistical methods like those used in combining forecasting techniques. Furthermore, weighting schemes can be constant over the time and reviewed at regular intervals or they can vary over the time according to a predefined pattern. Obviously, different weighting schemes can produce different patterns of the indicators which could even deliver contradictory signals.

Within the model based category, we can identify three main groups of methods: those based on linear models, on non-linear models and on pooling techniques. Considering linear models, dynamic factor models have emerged as the leading approach for the compilation of CCIs, sometimes in combination with other kind of models such as VAR, regression models and unobserved components ones. Concerning non-linear models, binary regression ones (e.g. LOGIT, PROBIT, etc.) and non-linear time series models (especially those allowing for the use of a treshold such as MS and SETAR models) have been the most widely used, even if other non-linear specification have received some attention especially in the recent years. An overview of non-linear models for compiling CCIs is provided by Ferrara and Mazzi (chapter 13). Concerning other non-linear techniques, it is worth mentioning the use of neural networks techniques Jagric (2003),Qi (2001). Unfortunately, computational problems as well as the relatively small number of observations characterising macroeconomic time series, have penalised an extensive use of such methods especially for the regular production of CCIs. Both for the dynamic factor models and the Markov switching models, there is a single unobservable force underlying the current and the future status of the economy, but in the former approach this is a continuous variable, while in the latter it is a discrete variable that evolves according to a Markov chain. This distinction makes dynamic factor models more adequate for compiling cyclical composite indicators, targeting the cyclical movement of the economy and its short-term evolution while the Markov switching ones are the most appropriate to compile indicators aiming at turning point detection (see chapter 2). Factor models provide a formalization of Burns and Mitchell (1946) notion of business cycles as co-movements in several variables. Leading references in the context of CCIs are Stock and Watson (1989),Stock and Watson (1991),Stock and Watson (1992), Forni et al. (2001), Altissimo et al. (2001), Altissimo et al. (2010). A detailed presentation of the Stock and Watson approach is provided by Moauro (chapter 11).

More recent work has focused on the construction of CCIs using mixed frequency factor models, in order to combine quarterly GDP data with timely monthly or even higher frequency information when available, see e.g. Mariano and Murasawa (2003), Frale et al. (2010) and Frale et al. (2011), Camacho and Perez-Quiros (2010).

Markov switching (MS) models formalize Burns and Mitchell (1946) notion that expansions and recessions are asymmetric. After the pioneering article by Hamilton (1989), vast literature followed, extending the basic model into several dimensions, e.g. Krolzig et al. (2002) consider multivariate MS error correction models, Diebold et al. (1994) and Filardo (1994) allow the transition probabilities to depend on exogenous variables, while Diebold and Rudebusch (1996), Kim and Nelson (1998), Filardo and Gordon (1999), Chauvet (1998) and Camacho et al. (2012) combine the characteristics of factor models and MS models by allowing MS features in the evolution of the factors.

Furthermore, Anas, Billio, Ferrara, Mazzi (chapter 14) developed multivariate Markov switching VAR model (MS VAR) to jointly derive turning points indicators for the business cycle and the growth cycle, which was consistent with the so called ABCD approach. Finally,Billio et al. (2013)have compared, in real-time, Markov switching and SETAR models to detect business cycle turning points.

If we consider not only the compilation methods but also the variable selection ones, then the distinction presented in this subsection becomes much less clearer. When considering also the variable selection methods, the distinction should be made more appropriately between non-parametric and parametric methods, instead of aggregation and model based ones. The borderline between the two groups of CCIs will be much less clear due to the fact that, unnecessarily, parametric or model based compilation methods are associated to parametric variable selection techniques and vice-versa. As an example, in Charpin et al. (2008) or Charpin and Mazzi (2010), non-parametric variable selection techniques, based on the LARS algorithm, are used to identify the most suitable variable to be introduced in the compilation framework based on a regression model. Furthermore, the variety of variable selection techniques and the difficulty of characterising them always as, either, parametric or non-parametric makes very complex and not necessarily useful to extend the classification proposed in this subsection, also to the variable selection step. Several aspects related to the complexity of the variable selection techniques are described in detail in chapter 6.

3.2.4 More ways to classify indicators

Beside the most commonly used ways to classify CCIs, we would just like to mention a couple of other classification criteria which have a certain interest, especially when trying to interpret the CCIs signals. The first classification criterion is represented by the kind of reference variable for the CCI. We assume that reference variables can be either directly observed (i.e. statistical indicators) or typically latent. If the Conference Board approach is a typical example of CCIs with a latent reference variable, the OECD CLI employs the GDP reference series. The second criterion is based on the typology of the data source used in compiling CCIs. Cyclical composite indicators can be typically based on qualitative information only, on both qualitative and quantitative information, and they can include, or not include, financial variables. Less often they can be based on quantitative statistical information only. The economic sentiment indicators released by DG ECFIN and the climate indicator released by the IFO Institute for Economic Research are CCIs based on qualitative information only, while the OECD CLI and the Conference Board include both qualitative and quantitative data. The rationale behind the use of different data sources in CCI's is based on some peculiarities of various data categories. Traditionally, qualitative data is considered very useful in producing short-term forecasts of key macroeconomic variables, even if they do not show any trend component. On the other hand, it provides a very good description of the growth cycle plus an irregular component, the size of which depends mainly on sampling aspects rather than intrinsic characteristic of the surveys. In most cases survey data can be even smoother than corresponding quantitative one (i.e. retail trade turnover). By contrast they do not provide a good representation of the business cycle due to their theoretical stationarity. Quantitative statistical data provides a better description of the classical cycle but they need to be de-trended to show growth cycle features. Finally, financial data is traditionally known for its leading features, both, in anticipating short-term macroeconomic movements and cyclical fluctuations. Nevertheless, the leading power of financial data can vary considerably across countries and economic regions.

3.2.5 Summary of Indicators

In this subsection we provide an overview of some of the most common CCIs, based on the taxonomy introduced so far and summarized in table 3.1.

Table 3.1: Taxonomy of selected indicators

	Non parametric	**Parametric - linear model**	**Parametric - nonlinear model**
Cyclical	ESI, PMI, IFO	BCI(DG ECFIN), CI(Conference Board), DZ-Euroland-Indicator, OECD CLI	
Turning Point	Heat maps (e.g. NBER)		BCCI, GCCI, ACCI, IESR, IARC
Short term growth		€-COIN EUROFRAME	IRC

IESR: conceived by the institute COE-REXECODE. It is a coincident indicator able to detect in real time peaks and troughs of the business cycle. It is based on the Markov-Switching model applied to various economic time series. The filtered probabilities are combined by taking into account the risks of false signals in order to provide a recession probability (a value of the indicator higher than 0.5).

IARC: prepared by the institute COE-REXECODE (Paris). It is a monthly indicator aimed at forecasting turning points in the growth cycle. Each month, the probability that each selected series has crossed a turning point is computed. Probabilities are aggregated by a weighting method to give an overall signal of a possible

downswing in the next nine months (a value of the indicator higher than 60 and lower than 80), or a very probable downswing in the three coming months (a value higher than 80).

IRC: coincident indicator of the business cycle available in real time, produced by the French COE-REXECODE institute. Through the use of temporal disaggregation method, using EU business surveys to extract a high frequency series (monthly GDP) from a low frequency series (quarterly GDP), the indicator provides monthly estimates of the quarterly 'underlying' growth of euro area GDP. The indicator relative to a calendar quarter is equal to the IRC of the last month of the quarter.

€-COIN: a real-time monthly estimate of euro area GDP growth, computed each month by the Banca d'Italia. It is obtained by collecting a large set of statistics and extracting information relevant to forecast future GDP. It tracks GDP growth by anticipating official GDP releases, by several months, by giving each month an early estimate of euro area growth in terms of quarter-on-quarter changes in GDP. Moreover, it sheds light on the underlying trend by removing short-run fluctuations and measurement errors from the growth rate; in this respect it is not only a forecast, but also an indicator of the true growth momentum in the euro area.

The EUROFRAME Euro Growth Indicator: calculated by the OFCE (Paris) in cooperation with the EU-ROFRAME group, which consists of: CPB (Den Haag), DIW (Berlin), ESRI (Dublin), ETLA (Helsinki), IFW (Kiel), NIESR (London), PROMETEIA (Bologna), WIFO (Vienna), and CASE (Poland). The purpose of this leading indicator is to anticipate the development of the GDP in the euro area two quarters ahead of official statistics. The indicator considers surveys from industry, construction, households, ISM survey of US industry, real euro/dollar exchange rate and European stock index. The indicator is based on ordinary least squares (OLS) estimation of the GDP growth rate, and released on a monthly basis.

The OECD Composite Leading Indicator: designed to predict cyclical turning points (peaks and troughs) in GDP as a proxy of the business growth cycle of the overall economy. The CLIs are composed of a set of economic indicators that jointly provide robust and early signals of future turning points in the economic activity. CLI indicators are pre-selected in order to cover, as far as possible, the key sectors of the economy as well as measure early stages of production; respond rapidly to changes in economic activity; and be sensitive to expectations of future activity.

IFO: Economic Climate Indicator designed to give an accurate picture of the economic situation and forecasts for economies on a quarterly basis. It consists of qualitative information: appraisals and expectations of economic experts. For the euro area the trend of the indicator correlates well with the actual business-cycle trend.

ESI: Economic Sentiment Indicator based on five qualitative surveys conducted on a monthly basis in the following areas: manufacturing industry, construction, consumers, retail trade and services.

PMI: Purchasing Managers' Indexes economic indicators derived from monthly surveys of private sector companies. The two principal producers of PMIs are Markit Group, which conducts PMIs for over 30 countries worldwide, and the Institute for Supply Management (ISM), which conducts PMIs for the US.

DG ECFIN Business Climate Indicator: designed to deliver a clear and early assessment of the cyclical situation of the euro area. The indicator uses, as input, five balances of opinions from DG ECFIN's Business and Consumer Surveys. The resulting 'common factor' may be read as a survey result: the higher the level, the healthier the cyclical situation and a rise in the indicator points to an upswing in activity and an improvement in business climate.

DZ-Euroland-Indicator: calculated by DZ Bank (monthly), aims at predicting recessions and has also proved to reflect movements in GDP growth rate. It is built on concepts implemented by the US Conference Board for monitoring cyclical indicators. Nine series for euro area or for the main countries are selected for their leading properties and aggregated to give a composite index which can be compared with the level of GDP in order to anticipate recessions.

BCCI, GCCI and **ACCI**: coincident Turning Point indicators of classical, growth and growth rate cycle respectively produced by Eurostat based on the Markov Switching model approach.

Heat Maps: a methodology analysis of a substantial number of series e.g.270 for NBER.Constructing a cyclical composite indicator.

3.3 Constructing a cyclical composite indicator

In this section we briefly review the main methods for the construction of a cyclical composite indicator (CCI). We start with the non-model based procedures and then discuss, in turn, factor based CCIs, Markov switching based CCIs, and pooling based CCIs. We concentrate our attention on coincident indicators but similar techniques can be generalised to the construction of leading and lagging indicators. Additional details and references can be found, e.g., in Marcellino (2006).

3.3.1 Aggregation based composite coincident indicators

In this case, the single components of an indicator are first transformed to have similar ranges and then aggregated using equal or different weights. A clear illustration is provided by (a slightly simplified version of) the step-wise procedure implemented by the Conference Board (CB), see http://www.conference-board.org and A. Ozyildirim (chapter 8) for details, which we will use as a benchmark for comparison with more sophisticated methods.

First, for each individual indicator, x_{it}, month-to-month symmetric percentage changes (spc) are computed as $x_{it\_spc} = 200 * (x_{it} - x_{it-1})/(x_{it} + x_{it+1})$. Second, for each x_{it_spc} a volatility measure, v_i, is computed as the inverse of its standard deviation. Third, each x_{it_spc} is adjusted to equalize the volatility of the components, the standardization factor being $s_i = v_i / \sum_i v_i$. Fourth, the standardized components, $m_{it} = s_i x_{it\_spc}$, are summed together with equal weights, yielding $m_t = \sum_i m_{it}$. Fifth, the composite indicator in levels is computed as

$$CCI_t = CCI_{t-1} * (200 + m_t)/(200 - m_t) \tag{3.1}$$

with the starting condition

$$CCI_1 = (200 + m_1)/(200 - m_1).$$

Finally, rebasing CCI to average 100 in 1996 yields the CCI_{CB}.

3.3.2 Factor-based composite indicators

A dynamic factor model was used to extract a coincident indicator by Stock and Watson (1989), with subsequent refinements of the methodology in Stock and Watson (1991), Stock and Watson (1992). The rationale of the approach is that all the coincident indicators are driven by a common force, the CCI, and by idiosyncratic components that are either uncorrelated across the variables under analysis or in any case common to only a limited subset of them. Hence, this approach formalizes Burns and Mitchell (1946) notion that business cycles represent co-movements in a set of series.

The particular model that SW adopted is the following,

$$\Delta x_t = \beta + \gamma(L)\Delta C_t + u_t \tag{3.2}$$

$$D(L)u_t = e_t \tag{3.3}$$

$$\phi(L)\Delta C_t = \delta + v_t \tag{3.4}$$

where x_t includes the components of the CCI, C_t is the single factor driving all variables, while u_t is the idiosyncratic component; Δ indicates the first difference operator, L is the lag operator and $\gamma(L), D(L), \phi(L)$ are, respectively, vector, matrix and scalar lag polynomials. SW used first differenced variables since unit root

tests indicated that the coincident variables were integrated, but not cointegrated. The model is identified by assuming that $D(L)$ is diagonal and e_t and v_t are mutually and serially uncorrelated at all leads and lags, which ensures that the common and the idiosyncratic components are uncorrelated. Moreover, ΔC_t should affect contemporaneously at least one coincident variable. Notice that the hypothesis of one factor, ΔC_t, does not mean that there is a unique source of aggregate fluctuations, but rather that different shocks have proportional dynamic effects on the variables.

For estimation, the model in (3.2)-(3.4) is augmented by the identity

$$C_{t-1} = \Delta C_{t-1} + C_{t-2}, \tag{3.5}$$

and cast into state-space form. The Kalman filter can then be used to write down the likelihood function, which is in turn maximized to obtain parameter and factor estimates. All the details are presented in Stock and Watson (1991).

A few additional comments are in order. First, the cyclical composite indicator, CCI_{SWt}, is obtained through the Kalman filter as the minimum mean squared error linear estimator of C_t using information on the coincident variables up to period t. Hence, the procedure can be implemented in real time, conditional on the availability of data on the coincident variables. By using the Kalman smoother rather than the filter, it is possible to obtain end of period estimates of the state of the economy, i.e., $C_{t|T}$. Second, it is possible to obtain a direct measure of the contribution of each coincident variable in x_t to the composite indicator by computing the response of the latter to a unit impulse in the former. Third, since data on some coincident variables are published with delay, they can be treated as missing observations and estimated within the state-space framework. Moreover, the possibility of measurement error in the first releases of the coincident variables can be also taken into consideration by adding an error term to the measurement equation. This is an important feature since data revisions are frequent and can be substantial, as for example testified by the revised US GDP growth rate data for 2001. Fourth, a particular time varying pattern in the parameters of the lag polynomials $D(L)$ and $\phi(L)$ can be allowed by using a time-varying transition matrix. Fifth, standard errors around the composite coincident indicator can be computed, even though they were not reported by SW. Sixth, a second factor can be included in the model, for the example to capture the behaviour of survey variables, given that the dynamics of soft data can be different from that of hard indicators, see e.g. Frale et al. (2010). Finally, the procedure can be extended to handle jointly quarterly and monthly (or higher frequency) indicators, see e.g. Mariano and Murasawa (2003), Frale et al. (2010) and Frale et al. (2011), Camacho and Perez-Quiros (2010).

A possible drawback of SW's procedure is that it requires an ex-ante classification of variables into coincident and leading or lagging, even though this is common practice in this literature, and it cannot be directly extended to analyze large datasets because of computational problems. Forni et al. (2000), Forni et al. (2001) (FHLR henceforth) proposed an alternative factor based methodology that addresses both issues and applied it to the derivation of a composite coincident indicator for the Euro area. They analyzed a large set of macroeconomic time series for each country of the Euro area using a dynamic factor model and decomposed each time series into a common, and an idiosyncratic, component where the former is the part of the variable explained by common Euro area shocks, the latter by variable specific shocks. The CCI_{FHLR} is obtained as a weighted average of the common components of the interpolated monthly GDP series for each country, where the weights are proportional to GDP and takes into account both within and across-countries cross correlations.

More specifically, the model FHLR adopted is

$$x_{it} = b_i'(L)v_t + \xi_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{3.6}$$

where x_{it} is a stationary univariate random variable, v_t is a $q \times 1$ vector of common shocks, $\chi_{it} = x_{it} - \xi_{it}$ is the common component of x_{it}, and ξ_{it} is its idiosyncratic component. The shock v_t is an orthonormal white noise process, so that $var(v_{jt}) = 1$, $cov(v_t, v_{t-k}) = 0$, and $cov(v_{jt}, v_{st-k}) = 0$ for any $j \neq s$, t and k. $\xi_N = \{\xi_{1t}, ..., \xi_{Nt}\}'$ is a wide sense stationary process and $cov(\xi_{jt}, v_{st-k}) = 0$ for any j, s, t and k. $b_i(L)$ is a $q \times 1$ vector of square summable, bilateral filters, for any i. Notice that SW's factor model (3.2) is obtained as

a particular case of (3.6) when there is one common shock ($q = 1$), $b_i(L) = \gamma_i(L)/\phi(L)$, and the idiosyncratic components are assumed to be orthogonal.

Grouping the variables into $x_{Nt} = \{x_{1t}, ..., x_{Nt}\}'$, FHLR also required x_{Nt} (and χ_{Nt}, ξ_{Nt} that are similarly defined) to have rational spectral density matrices, Σ_N^x, Σ_N^χ, and Σ_N^ξ, respectively. To achieve identification, they assumed that the first (largest) idiosyncratic dynamic eigenvalue, λ_{N1}^ξ, is uniformly bounded, and that the first (largest) q common dynamic eigenvalues, $\lambda_{N1}^\chi, ..., \lambda_{Nq}^\chi$, diverge, where dynamic eigenvalues are the eigenvalues of the spectral density matrix, see e.g. Brillinger (1981) (chapter 10). In other words, the former condition limits the effects of ξ_{it} on other cross-sectional units. The latter, instead, requires v_t to affect infinitely many units.

Let us assume for the moment that the number of common shocks is known. Then, FHLR suggested to estimate the common component of χ_{it} with the following step-wise procedure.

(i) Estimate the spectral density matrix of x_{Nt} as

$$\Sigma_N^T(\theta_h) = \sum_{k=-M}^{M} \Gamma_{Nk}^T \omega_k e^{-ik\theta_h}, \quad \theta_h = 2\pi h/(2M+1), \quad h = 0, ..., 2M, \tag{3.7}$$

where Γ_{Nk}^T is the sample covariance matrix of x_{Nt} and x_{Nt-k}, ω_k is the Bartlett lag window of size M ($\omega_k = 1 - k/(M+1)$), and M diverges but M/T tends to zero.

(ii) Calculate the first q eigenvectors of $\Sigma_N^T(\theta_h)$, $p_{Nj}^T(\theta_h)$, and the associated eigenvalues $\lambda_{j\theta}^x$, $j = 1, ..., q$, for $h = 0, ..., 2M$.

(iii) Define $p_{Nj}^T(L)$ as

$$p_{Nj}^T(L) = \sum_{k=-M}^{M} p_{Nj,k}^T L^k, \quad p_{Nj,k}^T = \frac{1}{2M+1} \sum_{h=0}^{2M} p_{Nj}^T(\theta_h) e^{ik\theta_h}, \quad k = -M, ..., M. \tag{3.8}$$

$p_{Nj}^T(L)x_{Nt}$, $j = 1, .., q$, are the first q estimated dynamic principal components of x_{Nt}.

(iv) The estimated common component of x_{it}, $\widehat{\chi}_{it}$, is the projection of x_{it} on present, past, and future dynamic principal components. FHLR proved that, under mild conditions, $\widehat{\chi}_{it}$ is a consistent estimator of χ_{it} when N and T diverge. Once the common component is estimated, the idiosyncratic one is obtained simply as a residual, namely, $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}_{it}$. Therefore, each variable can be decomposed into

$$x_{it} = \widehat{\chi}_{it} + \widehat{\xi}_{it}. \tag{3.9}$$

In practice, M and the number of leads (s) and lags (g) of $p_{Nj}^T(L)x_{Nt}$ to be included in the projection can be either chosen a priori or determined by minimizing the information criterion

$$\frac{T}{N} \sum_{i=1}^{N} \log \widehat{\sigma}_i + 2q(g+s+1), \tag{3.10}$$

where $\widehat{\sigma}_i$ is the estimated variance of $\widehat{\xi}_{it}$. Finally, FHLR suggested to determine the number of factors, q, on the basis of two properties: (a) the average over frequencies of the first q dynamic eigenvalues diverges, while the average of the $q+1^{th}$ does not; (b) there should be a big gap between the variance of x_{Nt} explained by the first q dynamic principal components and that explained by the $q+1^{th}$ principal component. An information criterion could be also used. In particular, the criterion that FHLR suggested for selection of g and s, equation (3.10) above, could be minimized also with respect to q.

A competing procedure for the analysis of dynamic factor models with a large number of variables was developed by Stock and Watson (2002a), Stock and Watson (2002b) (SW2 henceforth). The model by SW2, in its

time invariant formulation, can be written as

$$x_{nt} = \Lambda f_t + \xi_{nt}, \tag{3.11}$$

where f_t is an $r \times 1$ vector of common factors. Contrary to the specification by FHLR, the factors are not required to be uncorrelated in time, and they can be also correlated with the idiosyncratic component, only $var(f_t) = I$ is imposed for identification. Precise moment conditions on f_t and ξ_{nt}, and requirements on the loadings are given in SW.

The specification in (3.11) is related to the one by FHLR in (3.6). When $b_i(L)$ is unilateral and of finite order b, say $b_i(L) = b_{0i} - b_{1i}L - b_{bi}L^b$, the model in (3.6) can be written as in (3.11), where $f_t = (v_t, v_{t-1}, ..., v_{t-b})$ and the ith row of Λ has elements b_{0i}, b_{1i}, b_{bi}. Hence, $r = q(b+1)$, and the factors f_t are dynamically singular, i.e., the spectral density matrix of f_t has rank q.

To estimate the factors, SW2 define the estimators $\widehat{f}_t$ as the minimizers of the objective function

$$V_{nT}(f, \Lambda) = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \Lambda_i f_t)^2. \tag{3.12}$$

Under the hypothesis of k common factors, it turns out that the optimal estimators of the factors are ($\sqrt{T}$ times) the k eigenvectors corresponding to the k largest eigenvalues of the $T \times T$ matrix $n^{-1} \sum_{i=1}^{n} \underline{x}_i \underline{x}_i'$, where $\underline{x}_i = (x_{i1}, ..., x_{iT})$. These coincide with the principal components of the variables. Moreover, the optimal estimators of the loadings Λ are the OLS estimators of the coefficients in a regression of x_{it} on the k estimated factors $\widehat{f}_t$, $i = 1, ..., n$. Hence, a consistent estimator of the ith common component can be obtained as $\widehat{\chi}_{it} = \widehat{\Lambda}_i \widehat{f}_t$, and a natural choice for the estimator of the idiosyncratic component is $\widehat{\xi}_{it} = x_{it} - \widehat{\chi}_{it}$.

A convenient feature of the SW2 approach is that no future information is used for factor estimation, contrary to FHLR, and therefore the method can be applied in real time. The CCI_{SW} can be defined either as the single factor extracted from a set of coincident indicators or as an average of the common components of each single indicator. We will report results for the former, since it provides a direct generalization of the procedure in Stock and Watson (1989) and Watson (1989).

The methodology by FHLR was further refined by Altissimo et al. (2001), Altissimo et al. (2010) for real time implementation, and it is currently adopted to produce the CEPR's composite coincident indicator for the euro area, Eurocoin (see www.cepr.org). In particular, they exploited the large cross-sectional dimension for forecasting indicators available with delay and for filtering out high frequency dynamics. However, the main innovation is the use of an alternative estimator for the common components of the variables which does not require future information. The theory for the latter is presented in Forni et al. (2005) (FHLR2 henceforth).

While the analytical derivation of the method in FHLR2 is fairly complicated, its practical implementation is relatively easy. Let us reconsider the decomposition in (3.9), namely,

$$x_{it} = \widehat{\chi}_{it} + \widehat{\xi}_{it}, \tag{3.13}$$

and indicate the variance covariance matrix of $\widehat{\xi}_t$ by V_ζ.

Using, for example, the standard Choleski decomposition, it is possible to find a matrix P_ζ such that $P_\zeta V_\zeta P_\zeta' = I$. Multiplying both sides of (3.9) by P_ζ yields

$$P_\zeta x_{it} = P_\zeta \widehat{\chi}_{it} + P_\zeta \widehat{\xi}_{it} = \widehat{\alpha}_{it} + \widehat{\beta}_{it}, \tag{3.14}$$

where now the variance covariance of $\widehat{\beta}_t$ coincides with the identity matrix.

The principal components of $P_\zeta x_t$ are called generalized principal components of x_t by FHLR2, and the one-sided estimator of the common component is obtained by projecting the variables on the generalized principal components. Altissimo et al. (2001), Altissimo et al. (2010) construct Eurocoin as the weighted

average of common components of interpolated monthly GDP of euro area countries. For comparability with the other CCIs we have constructed, we will instead focus on the first generalized principal component of x_t, CCI_{FHLR2}.

3.3.3 Markov-switching-based CCI

The main criticism Sims (1989) raised in his comment to Stock and Watson (1989) is the use of a constant parameter model (even though, as remarked above, their framework is flexible enough to allow for parameter variation), and a similar critique can be addressed to the FHLR's method. Hamilton (1989) Markov switching model is a powerful response to this criticism, since it allows the growth rate of the variables (and possibly their dynamics) to depend on the status of the business cycle. A basic version of the model can be written as

$$\Delta x_t = c_{s_t} + A_{s_t} \Delta x_{t-1} + u_t, \tag{3.15}$$

$$u_t \sim i.i.d.N(0, \Sigma) \tag{3.16}$$

where, as in (3.2), x_t includes the coincident variables under analysis (or a single composite indicator), while s_t measures the status of the business cycle, with $s_t = 1$ in recessions and $s_t = 0$ in expansions, and both the deterministic component and the dynamics can change over different business cycle phases. The binary state variable s_t is not observable, but the values of the coincident indicators provide information on it.

With respect to the factor model based analysis, there is again a single unobservable force underlying the evolution of the indicators but, first, it is discrete rather than continuous and, second, it does not directly affect or summarize the variables but rather indirectly determines their behaviour that can change substantially over different phases of the cycle.

To close the model and estimate its parameters, an equation describing the behaviour of s_t is required, and it cannot be of autoregressive form as (3.4) since s_t is a binary variable. Hamilton (1989) proposed to adopt the Markov switching (MS) model, where

$$\Pr(s_t = j | s_{t-1} = i) = p_{ij}, \tag{3.17}$$

as previously considered by Lindgren (1978) and Neftci (1982) in simpler contexts. For expositional purposes we stick to the two states hypothesis, though there is some empirical evidence that three states can further improve the specification, representing recession, high growth and normal growth, see e.g. Kim and Murray (2002) for the US and Artis et al. (2004) for the Euro area. Further generalisations of this approach are presented in Ferrara, Mazzi (chapter 13) and Anas, Billio, Ferrara, Mazzi (chapter 14).

In our business cycle context, the quantity of special interest is an estimate of the unobservable current status of the economy and, assuming a mean square error loss function, the best estimator coincides with the conditional expectation of s_t given current and past information on x_t, which in turn is equivalent to the conditional probability

$$\zeta_{t|t} = \left(\begin{array}{c} \Pr(s_t = 0 | x_t, x_{t-1}, ..., x_1) \\ \Pr(s_t = 1 | x_t, x_{t-1}, ..., x_1) \end{array} \right). \tag{3.18}$$

Using simple probability rules, it follows that

$$\zeta_{t|t} = \left(\begin{array}{c} \frac{f(x_t|s_t=0,x_{t-1},...,x_1)\Pr(s_t=0|x_{t-1},...,x_1)}{f(x_t|x_{t-1},...,x_1)} \\ \frac{f(x_t|s_t=1,x_{t-1},...,x_1)\Pr(s_t=1|x_{t-1},...,x_1)}{f(x_t|x_{t-1},...,x_1)} \end{array} \right), \tag{3.19}$$

where

$$\Pr(s_t = i|x_{t-1}, ..., x_1) = \sum_{j=0}^{1} p_{ji} \Pr(s_{t-1} = j|x_{t-1}, ..., x_1), \qquad (3.20)$$

$$f(x_t|s_t = i, x_{t-1}, ..., x_1) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{T/2}} \exp[-(\Delta x_t - c_i - A_i \Delta x_{t-1})'\Sigma^{-1}(\Delta x_t - c_i - A_i \Delta x_{t-1})/2], \quad (3.21)$$

$$f(x_t, s_t = i|x_{t-1}, ..., x_1) = f(x_t|s_t = i, x_{t-1}, ..., x_1) \Pr(s_t = i|x_{t-1}, ..., x_1),$$

$$f(x_t|x_{t-1}, ..., x_1) = \sum_{j=0}^{1} f(x_t, s_t = j|x_{t-1}, ..., x_1), \quad i = 0, 1.$$

Hamilton (1994) or Krolzig (1997) provide additional details on these computations and formulae to calculate $\zeta_{t|T}$, i.e., the smoothed estimator of the probability of being in a given status in period t. Notice also that the first and last rows of (3.20) provide, respectively, the probability of the state and the density of the variables conditional on past information only.

The model in (3.15)-(3.17) can be extended in several dimensions, for example to allow for more states and co-integration among the variables, see e.g. Krolzig et al. (2002), or time-varying probabilities, as e.g. in Diebold et al. (1994) or Filardo (1994).

Factor models and Markov switching specifications capture two complementary and fundamental features of business cycles, namely, the diffusion of slow-down and recovery across many series and the different behavior of several indicators in expansions and recessions. They are not only flexible and powerful statistical tools but can be also given sound justifications from an economic theory point of view, see e.g. the overview in Diebold and Rudebusch (1996). The latter article represents also one of the earliest attempts to combine the two approaches by allowing the factor underlying SW's model to evolve, according to a Markov switching model. Yet, Diebold and Rudebusch (1996) did not jointly estimate the factor MS model. Such a task was tackled by Chauvet (1998) and Kim and Yoo (1995), using an approximated maximum likelihood procedure developed by Kim (1994) and by Kim and Nelson (1998) and Filardo and Gordon (1999) using Gibbs sampler techniques introduced by Albert and Chib (1993), Carter and Kohn (1994), and Shepard (1994). Camacho et al. (2012) present an interesting application to the construction of a CCI for the euro area.

3.3.4 Pooling-based CCI

Since the pioneering work of Bates and Granger (1969), it is well known that pooling can improve forecasting, i.e. estimating missing observations at the end of the sample, and there now exists a vast amount of empirical evidence to support their claim, see e.g. Timmermann (2005) for a recent overview. As discussed by Hendry and Clements (2004), possible reasons for the good performance of forecast pooling may be model mis-specification and parameter non-constancy that are attenuated by weighting. Marcellino (2007) shows that pooling is also quite effective for backdating and data interpolation, i.e. for estimating missing observations at the beginning or elsewhere in the sample.

Since constructing a CCI can be considered as a problem of estimation of missing observations (about the status of the economy), the cited evidence on the good performance of pooling suggests that combining a set of competing CCIs can improve upon the quality of each of the single CCIs.

The analysis is complicated by the fact that the forecasts can be compared with realized values after some time, while the status of the economy remains unobservable. This limits somewhat the range of feasible pooling techniques, as in the case of backdating or interpolation. Yet, simple combination methods such as averaging, possibly after trimming extreme values, work quite well in practice when compared with more sophisticated techniques, see e.g. Stock and Watson (1999) and Marcellino (2007).

Therefore, we will also experiment with averages of the single CCIs we described in the previous subsections, and we think that this is the first time that pooling is applied in the context of the construction of a composite coincident indicator, even though the construction of the non-model based indicators closely mimics pooling.

We will now apply the techniques reviewed so far for the construction of CCIs for the four largest countries in the euro area.

3.4 CCIs for European countries

We focus on the variables included into the CB composite coincident indicators for the European countries under analysis, in order to have a benchmark for the alternative CCI construction methods and avoid issues related with variable selection and transformation, even though the latter are very important, see e.g. Marcellino (2006) for details. While for some countries the Conference Board provides a longer time series for the CB indicators, there have been changes in the raw coincident variables used to compute the indicator. For this reason, we focus on the last 20 years of data, as in this period there were no changes in the underlying coincident variables. In particular, we have the following list.

- France. Sample: 1992Feb-2011nov
    - FRIP - Industrial Production
    - FRPCON - Personal Consumption
    - FREMPL - Numbers of Employees in the Private Non-Agricultural Sector
    - FRWAGE - Wage & Salaries Paid by Non-Financial & Financial Companies
- Germany. Sample: 1994Jan-2011nov
    - GEIP - Industrial Production
    - GEMP - Employment - Number of People Employed
    - GERS - Retail sales
    - GEMSA_MA3 - Manufacturing Sales
    - Spain. Sample: 1995Jan-2011nov
    - SPCONSP - Final Household Consumption (Q)
    - SPIPROD - Industrial Production Excluding Construction (3 month moving average)
    - SPRSMRKT - Retail Sales Survey (s.a.) 1/95 - present
    - SPEMP - Employment, Thousand, s.a.
- United Kingdom: Sample: 1992Feb-2011nov
    - UKIP - Industrial Production
    - UKRS - Retail Sales
    - UKEMP_M - Employment - LFS
    - UKHDINC_INT - Real Household Disposable Income

The (logs of the) variables for each country are graphed in figure 3.1, after normalization. Overall, the indicators for each country follow the same trend and their peaks and troughs structure at the quarterly level mimics that of GDP growth, as we will see later on. However, there are also evident differences in the behaviour of

the single variable for each country, which provides support for their combination into a CCI rather than for the selection of a single variable.

Figure 3.1: Components of the coincident indexes
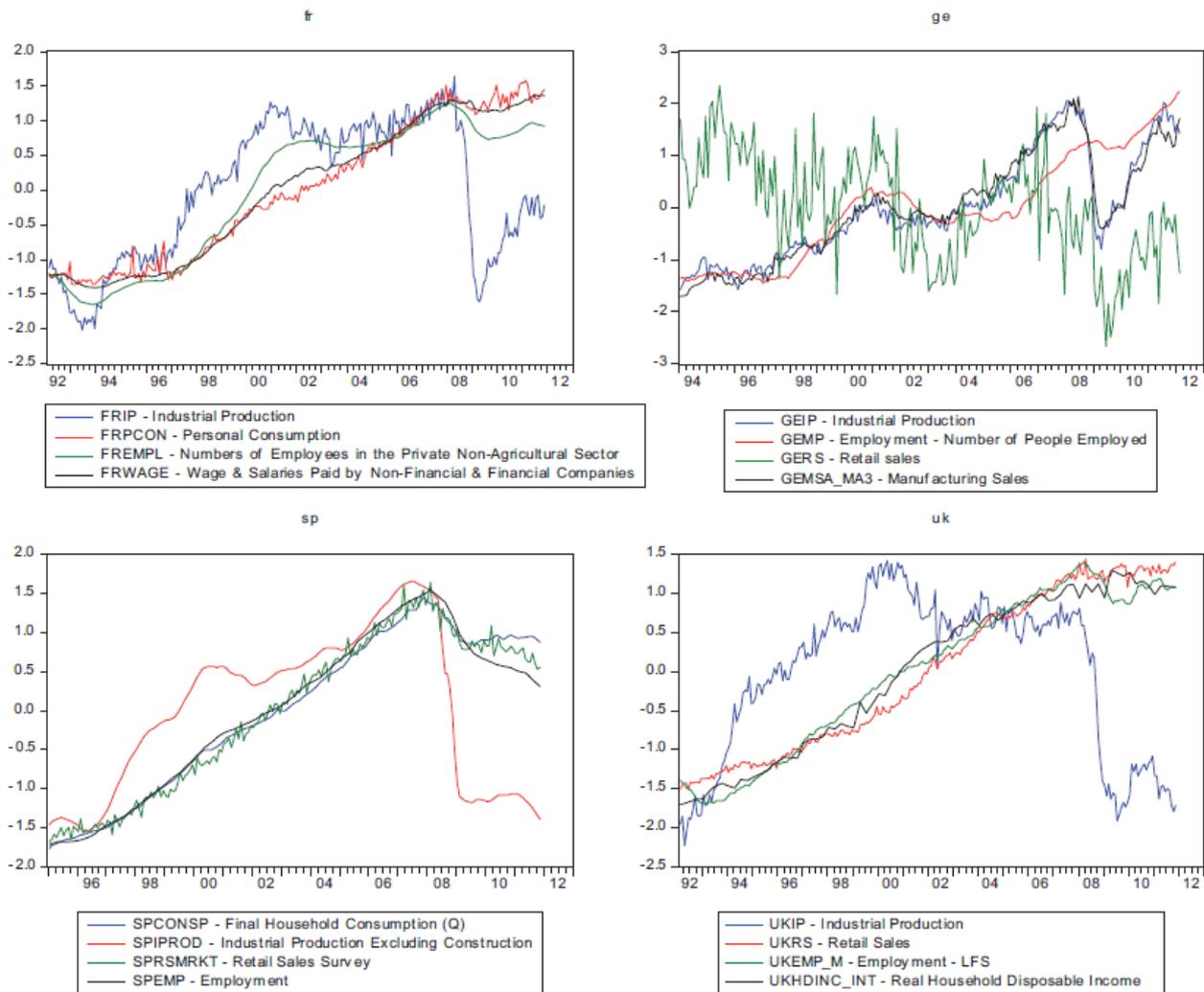
3.4.1 Factor-based CCIs

The factor-based methods for the construction of a CCI described in the previous Section require the input variables to be stationary. From the graphs in figure 3.1 the single coincident variables present either a trending behaviour or at least persistent deviations from the mean. These features are confirmed by unit root tests, which do not reject the null hypothesis of a unit root for any of the variables. Therefore, as SW and FHLR, we will work with the (month on month) growth rates of the single variables.[1]

To start with, we adopt a parametric factor model as in SW and we construct the CCI_{SW} as the (cumulated) estimated factor C_{ct}, $c =$ France, Germany, Italy, Spain. For each country, we consider the following

---

[1] All the factor based methods we have described do not take co-integration across the single indicators into account. The omission of an error correction term can be a serious issue, see e.g. Emerson and Hendry (1996) in a related context. Hence, in future research this extension could be considered.

specification, where the variables are the demeaned log differences of the levels:

$$
\begin{cases}
y_{1ct} = \gamma_{1c}\Delta C_{ct} + u_{1ct} \\
y_{2ct} = \gamma_{2c}\Delta C_{ct} + u_{2ct} \\
y_{3ct} = \gamma_{3c}\Delta C_{ct} + u_{3ct} \\
y_{4ct} = \gamma_{4c}\Delta C_{ct} + u_{4ct} \\
u_{ict} = \psi_{1ic}u_{ict-1} + \psi_{2ic}u_{ict-2} + \varepsilon_{ict}; \quad \varepsilon_{ict} \sim iidN(0,\sigma_{ic}^2); \quad i=1,2,3,4; \\
\Delta C_{ct} = \varphi_1\Delta C_{ct-1} + \varphi_2\Delta C_{ct-2} + v_{ct}; \quad v_{ct} \sim iidN(0,1) \\
COV(\varepsilon_{ict}, v_{cs}) = 0 \; \forall c, \; \forall i, \; \forall s, \; \forall t
\end{cases}
\tag{3.1}
$$

It is not guaranteed that this model will feature homoskedasticity, normality and lack of correlation of the disturbances ε_{ict}. A possible solution to address the partial misspecification of the parametric factor model in (3.1) is to resort to nonparametric techniques to estimate the common factor and obtain alternative factor-based CCIs. Though the methods by FHLR2 and SW2 are particularly suited when the number of variables under analysis is large, it is interesting to evaluate their performance in our context.

With reference to the models in the previous Section, for SW2 we use one factor and for FHLR2 we set the bandwidth parameter at M=12 and use one factor both in the first step (i.e. to compute the variance covariance matrix of the common components obtained using FHLR) and in the second step.

In figure 3.2 we graph the (standardized) Conference Board CCI and the three versions of the factor based CCIs, namely, SW, SW2 and FHLR2, and the CCI obtained by pooling all these CCIs. All composite indicators tend to move closely together. The visual impression is confirmed by the correlations reported in the panel A of tables 3.A.1-3.A.4.

Figure 3.2: Alternative CCIs for European countries - Levels
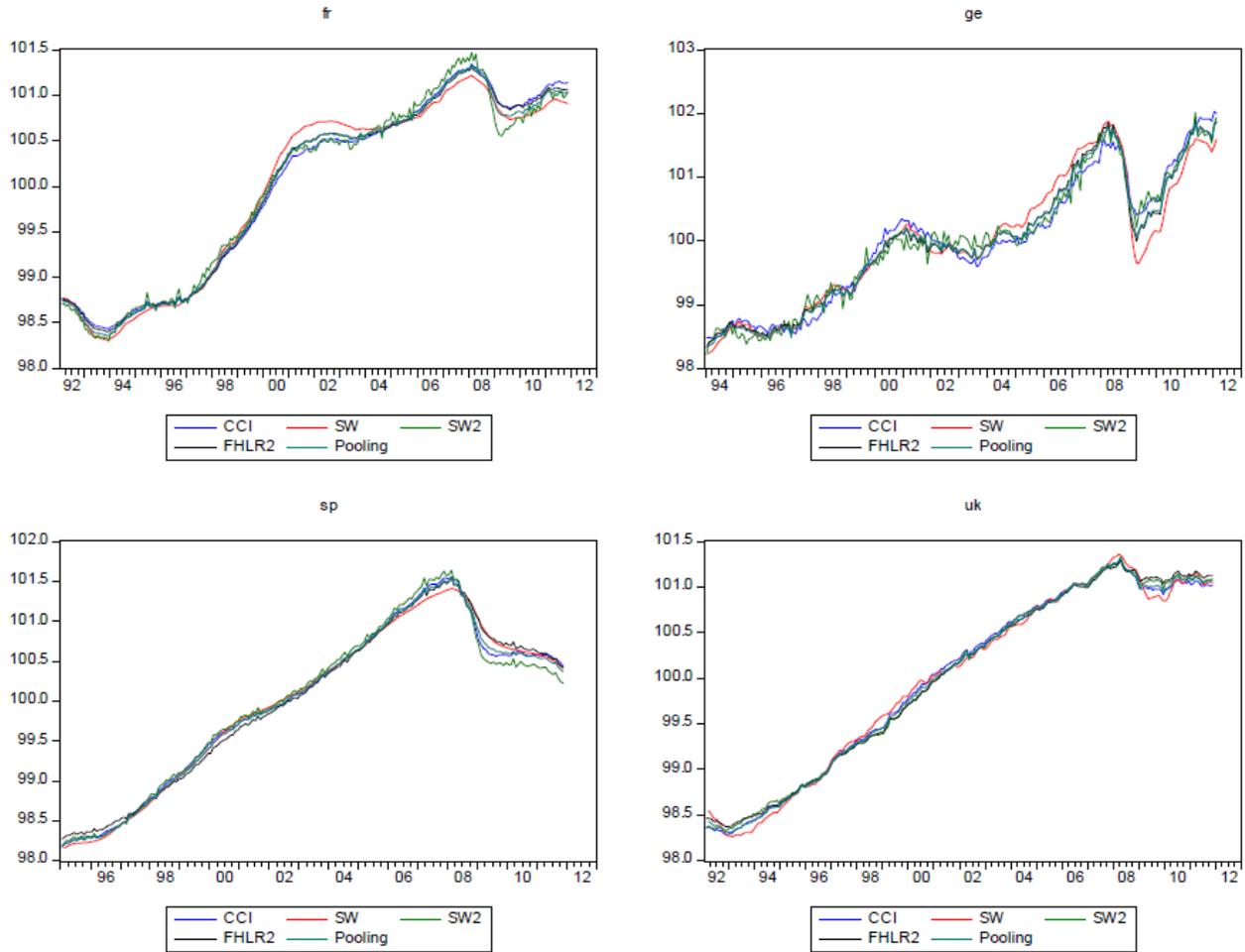
A similar finding emerges from the 6-month growth rates of the CCIs, Figure 3.3, and from the bandpass filtered CCIs where we apply the bandpass HP filter proposed in Artis et al. (2004) to emphasize the business cycle frequencies (between 1.5 and 8 years), Figure 3.4. The related correlations are reported in the panels B and C of tables 3.A.1-3.A.4. They remain quite high, with the lowest values typically achieved either by SW with either CB or FHLR2 (still these values are larger than 0.7), and figures typically higher than 0.90 for SW2 and FHLR2. The use of growth rates or filtered data also emphasizes the close similarity of the indexes at turning points.

Figure 3.3: Alternative CCIs for European countries - 6-month growth rates
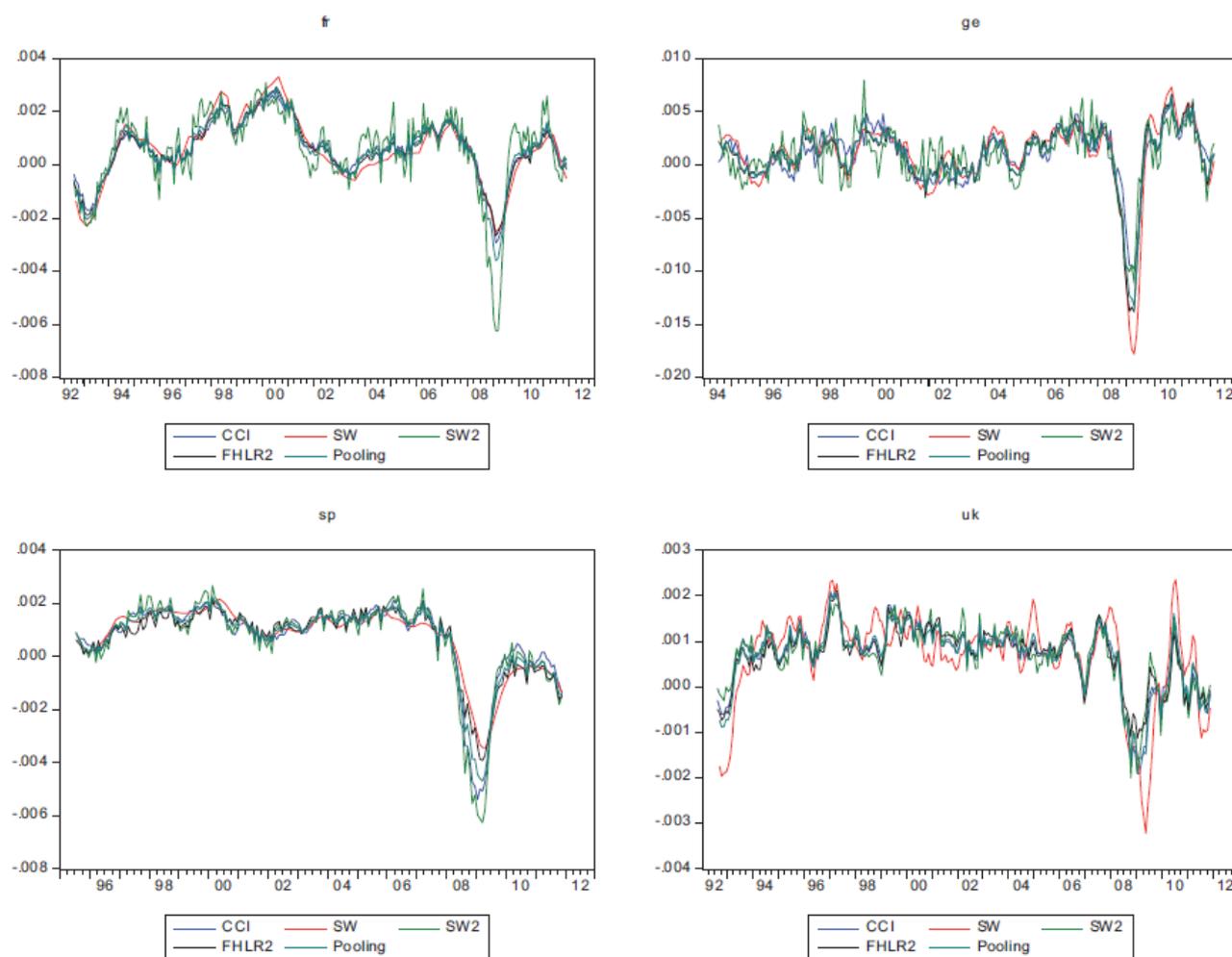
Figure 3.4: Alternative CCIs for European countries - HP bandpass filtered
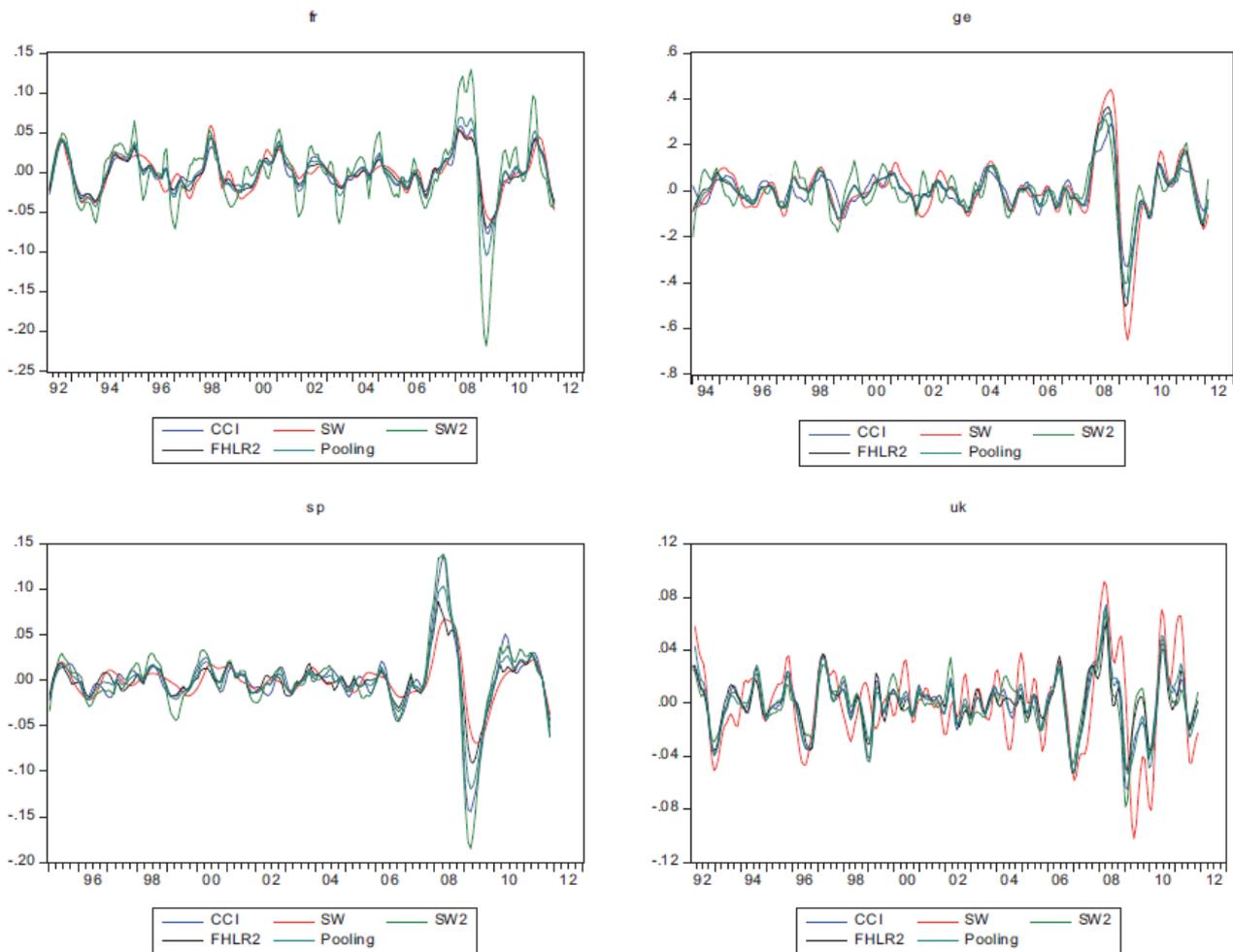
Finally, in Table 3.A.5 we report the correlations with yearly GDP growth of the various CCIs and of the Economic Sentiment Indicator (ESI), produced by the European Commission and based on surveys. Interestingly, for all countries, the ESI has slightly lower correlation with GDP growth than any other CCI under consideration. However, an important practical advantage of the survey-based ESI is its timely availability at the end of the reference month.

3.4.2 Markov-switching-based CCIs

To evaluate the usefulness of the Markov-switching approach for the construction of CCIs for European countries, we have estimated for each country the MS-VAR(1) model

$$\Delta x_t = c_{s_t} + A_{s_t}\Delta x_{t-1} + u_t, \quad u_t \sim i.i.d. N(0, \Sigma), \tag{3.2}$$

where s_t is the binary expansion / recession indicator. More complicated specifications, with additional lags or switches in the variances, are typically hardly estimable in our context, due to the rather short sample available and the limited number of cycles.

Figure 3.5: MS-CCI for European countries, filtered probabilities, and 6-month growth rate in the CCI_{CB}.
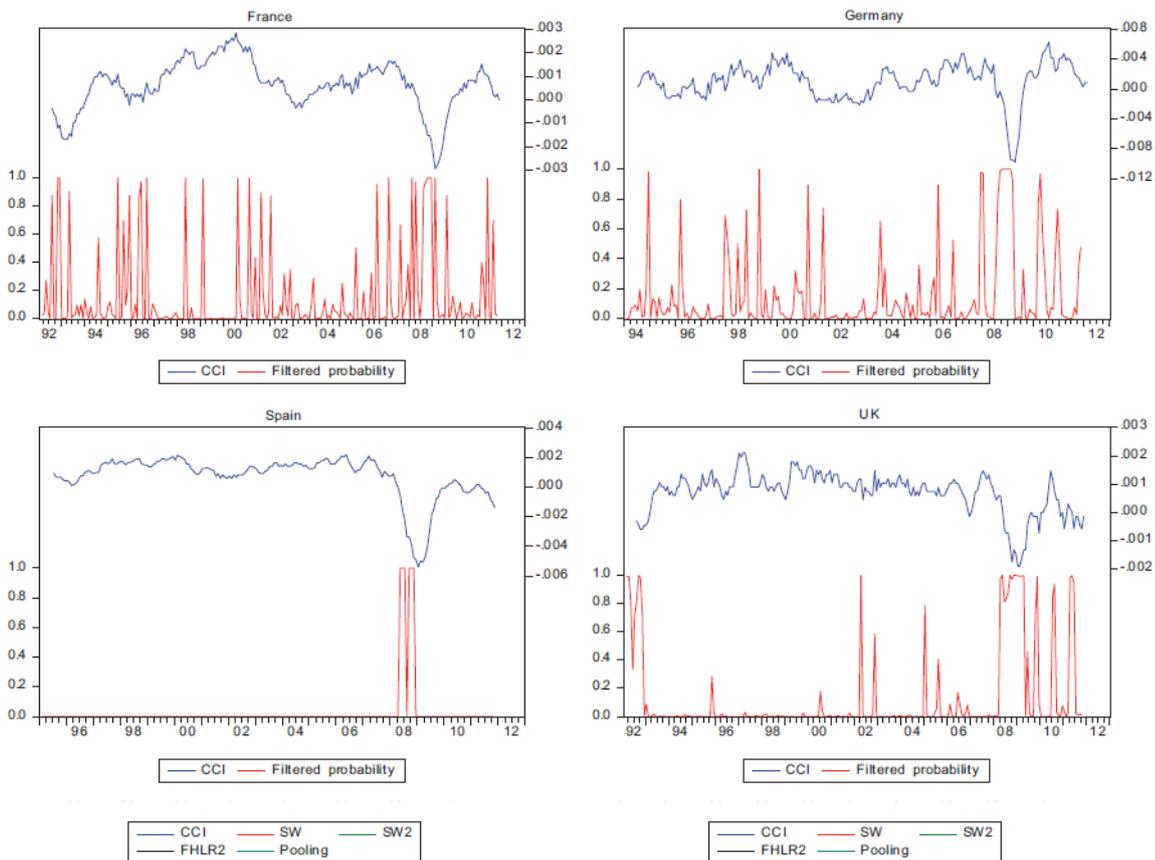
Figure 3.6: MS-CCI for European countries, smoothed probabilities, and 6-month growth rate in the CCI_{CB}.
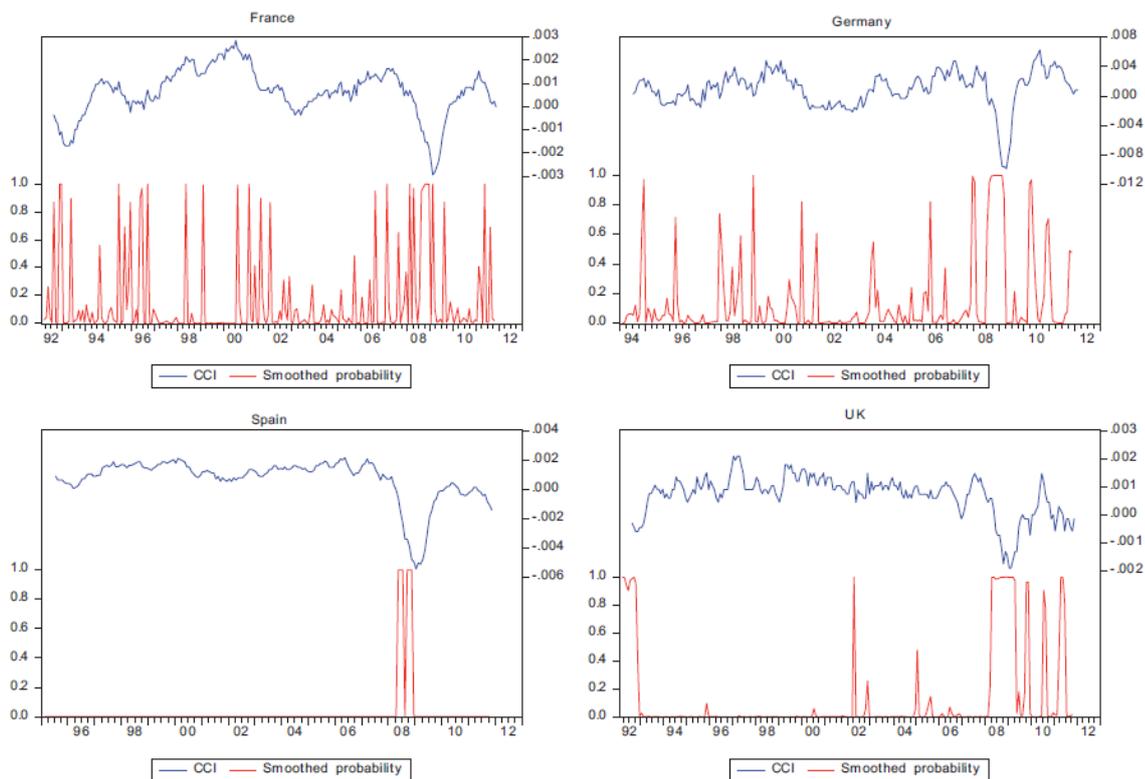
In figure 3.5 we compare the filtered probability of recession resulting from the MS-VAR with the 6 month growth rate in the CCI_{CB}. We would expect higher probability of recessions associated with marked slow-downs in the growth of the CCI_{CB}. This appears to be the case for the UK and Spain, where the major recessions, as dated by Artis et al. (2004), are associated with increases in the probability of recession. However, the picture is different for France and Germany, where the probability of recession is too high too frequently. Similar results are obtained with the smoothed probabilities, which are based on the full sample rather than recursively updated, see Figure 3.6. Overall, at least for the sample period and time series we consider, there seem to be minor or no gains from the construction of MS based CCIs with respect to the other ones.

3.5 Conclusions

In this paper, we have first provided some relevant definitions concerning cyclical composite indicators and we have developed a tentative taxonomy for them. Then, we have surveyed a variety of statistical methods for the construction of cyclical composite indicators, and proposed a novel pooling-based procedure. We have then applied some of the techniques to construct CCIs to monitor economic conditions in some of the largest European countries.

Simple non-model based CCIs for the European countries, which are averages of standardized selected single coincident variables, yield in general similar results as more complicated methods. However, the more sophisticated model based methods can provide a statistical framework for the computation of standard errors around the CCI, the unified treatment of data revisions and missing observations, and the development of composite leading indicators (see e.g. Marcellino (2006)).

Among the model based approaches to CCIs construction, factor based methods provide good results with limited estimation problems even with short time series. Markov switching methods are interesting but, with the short and noisy time series typically available for Europe, estimation can be a serious issue.

Pooling is not particularly helpful in this context, likely due to the high correlation across the combined CCIs. It can improve the correlation with GDP growth for a few countries, but in general the gains are minor. On the other hand, the good performance of the CCIs_{CB} can be related to pooling, where single variables rather than composite indicators are combined, which increases their variability.

Finally, our rather simple CCIs perform well when compared with the Economic Sentiment Indicator, which is by the way more timely.

Overall, the results indicate that it is possible to achieve a substantial consensus on the current status of the economy, which is an important finding for economic policy and, more generally, for decision making.

Annex

3.A Correlations of alternative CCIs

The following tables displays the correlations of alternative CCIs. CCIs are: CB: Conference Board, SW: Stock and Watson (1989), FHLR2: Forni et al. (2005), SW2: Stock and Watson (2002a), Stock and Watson (2002b), POOL: pooling of all CCIs. RGDP is real GDP. Panels A, B, C are based on monthly data and contain respectively results for the level of CCIs, the 6 months %change in CCIs, and for the CCIs filtered as in Artis et al. (2004). Panels D and E are based on quarterly data and contain respectively results for the level of CCIs and the 2 quarters %change in CCIs. The sample is different and is presented under each of them.

Table 3.A.1: France, Correlations of alternative CCIs

Monthly data

A. Levels	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.992	0.995	0.999	0.998	
SW		1.000	0.992	0.996	0.997	
FHLR2			1.000	0.995	0.997	
SW2				1.000	1.000	
POOL					1.000	
B. 6 months % change	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.958	0.902	0.990	0.991	
SW		1.000	0.807	0.982	0.960	
FHLR2			1.000	0.876	0.935	
SW2				1.000	0.989	
POOL					1.000	
C. Filtered data	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.900	0.888	0.972	0.978	
SW		1.000	0.722	0.953	0.903	
FHLR2			1.000	0.864	0.943	
SW2				1.000	0.979	
POOL					1.000	

Quarterly data

D. Levels	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.992	0.995	0.999	0.998	0.992
SW		1	0.992	0.996	0.997	0.973
FHLR2			1	0.995	0.998	0.988
SW2				1	1.000	0.988
POOL					1	0.987
RGDP						1
E. 2 quarters % change	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.959	0.903	0.990	0.990	0.896
SW		1	0.813	0.978	0.960	0.830
FHLR2			1	0.887	0.938	0.916
SW2				1	0.990	0.868
POOL					1	0.910
RGDP						1

Sample is 1992:2 2011:11

Table 3.A.2: Germany, Correlations of alternative CCIs

Monthly data

A. Levels	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.961	0.980	0.987	0.990	
SW		1.000	0.965	0.988	0.986	
FHLR2			1.000	0.991	0.992	
SW2				1.000	0.999	
POOL					1.000	
B. 6 months % change	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.908	0.681	0.912	0.926	
SW		1.000	0.762	0.951	0.966	
FHLR2			1.000	0.892	0.882	
SW2				1.000	0.996	
POOL					1.000	
C. Filtered data	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.937	0.706	0.924	0.942	
SW		1.000	0.755	0.943	0.966	
FHLR2			1.000	0.904	0.882	
SW2				1.000	0.995	
POOL					1.000	

Quarterly data

D. Levels	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.961	0.978	0.987	0.990	0.982
SW		1	0.962	0.987	0.986	0.970
FHLR2			1	0.990	0.991	0.987
SW2				1	1.000	0.988
POOL					1	0.990
RGDP						1
E. 2 quarters % change	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.908	0.577	0.892	0.910	0.860
SW		1	0.669	0.938	0.955	0.900
FHLR2			1	0.852	0.836	0.583
SW2				1	0.996	0.854
POOL					1	0.867
RGDP						1

Sample is 1994:1-2011:11.

Table 3.A.3: Spain, Correlations of alternative CCIs

Monthly data

A. Levels	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.997	0.997	0.997	1.000	
SW		1.000	0.992	0.998	0.999	
FHLR2			1.000	0.992	0.997	
SW2				1.000	0.998	
POOL					1.000	
B. 6 months % change	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.924	0.984	0.951	0.986	
SW		1.000	0.909	0.967	0.963	
FHLR2			1.000	0.960	0.986	
SW2				1.000	0.987	
POOL					1.000	
C. Filtered data	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.819	0.969	0.922	0.979	
SW		1.000	0.796	0.913	0.892	
FHLR2			1.000	0.941	0.979	
SW2				1.000	0.977	
POOL					1.000	

Quarterly data

D. Levels	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.997	0.997	0.997	1.000	0.974
SW		1	0.992	0.998	0.999	0.980
FHLR2			1	0.992	0.997	0.955
SW2				1	0.998	0.983
POOL					1	0.975
RGDP						1
E. 2 quarters % change	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.924	0.986	0.953	0.987	0.905
SW		1	0.906	0.964	0.962	0.968
FHLR2			1	0.960	0.985	0.880
SW2				1	0.986	0.924
POOL					1	0.933
RGDP						1

Sample is 1995:1-2011:11.

Table 3.A.4: UK, Correlations of alternative CCIs

Monthly data

A. Levels	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.998	0.999	0.998	1.000	
SW		1.000	0.996	0.996	0.998	
FHLR2			1.000	1.000	0.999	
SW2				1.000	0.999	
POOL					1.000	
B. 6 months % change	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.834	0.952	0.960	0.984	
SW		1.000	0.689	0.798	0.902	
FHLR2			1.000	0.939	0.928	
SW2				1.000	0.967	
POOL					1.000	
C. Filtered data	CB	SW	FHLR2	SW2	POOL	
CB	1.000	0.740	0.917	0.943	0.973	
SW		1.000	0.506	0.662	0.852	
FHLR2			1.000	0.906	0.871	
SW2				1.000	0.938	
POOL					1.000	

Quarterly data

D. Levels	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.998	0.999	0.998	1.000	0.997
SW		1	0.996	0.996	0.998	0.996
FHLR2			1	1.000	1.000	0.998
SW2				1	0.999	0.997
POOL					1	0.998
RGDP						1
E. 2 quarters % change	NMB	SW	FHLR2	SW2	POOL	RGDP
NMB	1	0.855	0.950	0.964	0.986	0.791
SW		1	0.718	0.825	0.915	0.702
FHLR2			1	0.941	0.932	0.714
SW2				1	0.972	0.690
POOL					1	0.764
RGDP						1

Sample is 1992:1-2011:11.

Table 3.A.5: CCIs:correlations with individual countries 1-year GDP growth

ESI		**SW2**	
France	0.84	France	0.96
Germany	0.83	Germany	0.86
Spain	0.88	Spain	0.92
UK	0.80	UK	0.86
CB		**FHLR2**	
France	0.92	France	0.90
Germany	0.90	Germany	0.94
Spain	0.94	Spain	0.96
UK	0.90	UK	0.80
SW1		**Pooling**	
France	0.86	France	0.93
Germany	0.94	Germany	0.94
Spain	0.98	Spain	0.96
UK	0.81	UK	0.88

The table reports the correlations between the CCis and the annual GDP growth rate. CCIs are: CB Conference Board, SW: Stock and Watson (1989), FHLR2: Forni et al. (2005), SW2: Stock and Watson (2002a),Stock and Watson (2002b), POOL: pooling of all CCIs.

Bibliography

Albert J., Chib S. (1993), Bayesian analysis via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts, Journal of Business and Economic Statistics 11: 1-15.

Altissimo F., Bassanetti A., Cristadoro R., Forni M., Lippi M., Reichlin L., Veronese G. (2001), EuroCoin: A real time coincident indicator of the Euro area business cycle, CEPR Working Paper 3108.

Altissimo F., Riccardo C., Mario F., Marco L., Veronese G. (2010), New eurocoin: Tracking economic growth in real time. The Review of Economic and Statistics, 92, 1024–1034.

Anas J. and Ferrara L. (2002), Un indicateur d'entrée et sortie de récession : application aux Etats Unis, Document de travail No. 58, Centre d'Observation Economique.

Anas J., Billio M., Ferrara L., Mazzi G.-L. (2008), A system for dating and detecting turning points in the euro area, The Manchester School, Vol. 76, No. 5, pp. 549-577.

Angelini E. , Camba-Mendez G., Giannone D., Reichlin L., Rünstler G. (2011), Short-term forecasts of euro area GDP growth, Econometrics Journal, Royal Economic Society, vol. 14(1), pages C25-C44, February.

Artis M.J., Krolzig H.M., Toro J. (2004), The European business cycle, Oxford Economic Papers 56(1), 1-44.

Artis M.J., Marcellino M., Proietti T. (2004), Dating the Euro area business cycle, Oxford Bulletin of Economics and Statistics, 66, 537-565.

Banbura M, Giannone D, Reichlin L. (2011), Nowcasting. In Oxford Handbook on Economic Forecasting, Clements MP, Hendry DF (eds). Oxford University Press: Oxford. .

Banbura, M., Giannone D., Modugno M., Reichlin L. (2012a), Now-Casting and the Real-Time Data Flow, Working Papers ECARES 2012-026, ULB – Université Libre de Bruxelles.

Banbura M., Giannone D., Modugno M., Reichlin L. (2012b), Nowcasting with Daily Data, 2012 Meeting Papers 555, Society for Economic Dynamics.

Bates J.M., Granger C.W.J. (1969), The combination of forecasts, Operations Research Quarterly 20: 451-468.

Billio M., Ferrara L., Guégan D., Mazzi G.L. (2013), Evaluation of Regime Switching Models for Real-Time Business Cycle Analysis of the Euro Area, Journal of Forecasting, John Wiley & Sons, Ltd., vol. 32(7), pages 577-586, November.

Boldin M.D. (1994), Dating Turning Points in the Business Cycle, Journal of Business, Vol. 67 (1) pp. 97-131. University of Chicago Press.

Brillinger D.R. (1981), Time series data analysis and theory, Holt, Rinehart, and Winston (New York).

Burns A.F., Mitchell W.C. (1946), Measuring business cycles, NBER Studies in Business Cycles no. 2 (New York).

Camacho M., Perez-Quiros G. (2010), Introducing the EURO-STING: Short Term INdicator of Euro Area Growth, Journal of Applied Econometrics, 25: 663-694.

Camacho M., Perez-Quiros G., Poncela P. (2012), Markov-switching dynamic factor models in real time, Working paper n. 1205 at the Central Bank of Spain.

Camba-Méndez G., Kapetanios G. (2004), Forecasting euro area inflation using dynamic factor measures of underlying inflation, Journal of Forecasting, November 2005, Vol. 24, pp. 491-503.

Carriero A., Marcellino M. (2007), A comparison of methods for the construction of composite coincident and leading indexes for the UK, International Journal of Forecasting, 23, 219-236.

Carriero A., Marcellino M. (2011), Sectoral Survey-based Confidence Indicators for Europe, Oxford Bulletin of Economics and Statistics, 73, 175–206.

Carter C.K., Kohn R. (1994), On Gibbs sampling for state space models, Biometrika 81: 541-55.

Charpin F., Mazzi G.L. (2010), Early estimate of the quarterly GDP growth of euro area, 30th CIRET Conference, New York.

Charpin F., Mazzi G.L., Mathieu C. (2008), Construction of coincident indicators for the euro area, 5th EURO-STAT Colloquium on Modern Tools For Business Cycle Analysis, Luxembourg, 29th September – 1st October. 2008.

Charpin F., Mazzi G.L., Mathieu C. (2008), Construction of coincident indicators for euro area key macroeconomic variables, 28th International Symposium on Forecasting, Nice, June 23 2008.

Charpin F., Mazzi G.L., Moauro F. (2012), Early estimates of GDP growth rates by error correction models, 31st CIRET Conference, Vienna, 5 – 8 September, 2012.

Chauvet M., Hamilton J.D. (2006), Dating business cycle turning points, Contributions to Economic Analysis 276, 1-54.

Chauvet M. (1998), An econometric characterization of business cycle dynamics with factor structure and regime switching, International Economic Review 39(4): 969-996.

Chin D., Geweke J., Miller P. (2000), Predicting turning points, Federal Reserve Bank of Minneapolis Staff Report no. 267.

Diebold F.X., Lee J.H., Weinbach G.C. (1994), Regime switching with time-varying transition probabilities, in C. Hargreaves, ed., Non-stationary time-series analyses and cointegration, Oxford University Press (Oxford): 283-302.

Diebold F.X., Rudebusch G.D. (1996), Measuring business cycles: a modern perspective, The Review of Economics and Statistics 78(1): 67-77.

Emerson R.A., Hendry D.F. (1996), An evaluation of forecasting using leading indicators, Journal of Forecasting, 15, 271-291.

Filardo A.J. (1994), Business cycle phases and their transitional dynamics, Journal of Business and Economic Statistics 12(3):299-308.

Filardo A.J., Gordon S.F. (1999), Business cycle turning points: two empirical business cycle model approaches, in: P. Rothman, ed., Nonlinear time series analysis of economic and financial data, vol. 1 (Kluwer Academic Publishers), ch. 1: 1-32.

Forni M., Hallin M., Lippi M., Reichlin L. (2000), The generalized factor model: identification and estimation, The Review of Economics and Statistics 82(4): 540-554.

Forni M., Hallin M., Lippi M., Reichlin L. (2001), Coincident and leading indicators for the Euro area, The Economic Journal 111(May): C62-C85.

Forni M., Hallin M., Lippi M., Reichlin L. (2005), The generalized dynamic factor model: One-sided estimation and forecasting, Journal of the American Statistical Association, Vol 100 (pp. 830 – 840).

Frale C., Marcellino M., Mazzi G.L., Proietti, T. (2010), Survey data as coincident or leading indicators, Journal of Forecasting, 29, 109–131.

Frale C., Marcellino M., Mazzi G.L., Proietti T. (2011), EUROMIND: A monthly indicator of the euro area economic conditions, Journal of the Royal Statistical Society, A, 174, 439-470.

Hamilton J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57:357-384.

Hamilton J.D. (1994), Time Series Analysis, Princeton University Press (Princeton).

Harding D. (2010), Applying shape and phase restrictions in generalized dynamic categorical models of the business cycle, NCER Working Paper Series 58, National Centre for Econometric Research.

Harding D., Pagan A.R. (2011), Econometric Analysis and Prediction of Recurrent Events, No 75, NCER Working Paper Series from National Centre for Econometric Research.

Hendry D.F., Clements M.P. (2004), Pooling of Forecasts, Econometrics Journal, 7, 1-31.

Jagric T. (2003), Forecasting with Leading Economic Indicators — A Neural Network Approach, Operations Research Proceedings Vol. 2002.

Johansen S.(1988), Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control 12: 231-254.

Kim C.J. (1994), Dynamic linear models with Markov switching, Journal of Econometrics 60: 1-22.

Kim M.J., Yoo J.S. (1995), New index of coincident indicators: a multivariate Markov switching factor model approach, Journal of Monetary Economics 36: 607-630.

Kim C.J., Nelson C. R. (1998), Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, The Review of Economics and Statistics 80: 188-201.

Kim C.J., Murray C. (2002), Permanent and transitory components of recessions, Empirical Economics 27: 163-183.

Krolzig H.M. (1997), Markov switching vector autoregressions, Modelling, statistical inference and application to business cycle analysis, Springer (Berlin).

Krolzig H.M., Marcellino M. and Mizon (2002), A Markov-switching vector equilibrium correction model of the UK labour market, Empirical Economics 27(2): 233-254.

Lindgren G. (1978), Markov regime models for mixed distributions and switching regressions, Scandinavian Journal of Statistics 5: 81-91.

Marcellino M. (2004), Forecast pooling for short time series of macroeconomic variables, Oxford Bulletin of Economics and Statistics, 66, 91-112.

Marcellino M. (2006), Leading indicators, in Elliott, G., Granger, C.W.J. and Timmermann, A. (eds), Handbook of Economic Forecasts, Amsterdam: Elsevier.

Marcellino M. (2007), Pooling based interpolation and back-dating, Journal of Time Series Analysis, 28, 53-71.

Marcellino M., Stock J.H., Watson M.W. (2003), Macroeconomic forecasting in the Euro area: country specific versus Euro wide information, European Economic Review, 47, 1-18.

Mariano R.S., Murasawa Y. (2003), A new coincident index of business cycles based on monthly and quarterly series, Journal of Applied Econometrics, 18, 427–443.

Moore G.H., Shiskin J. (1967), Indicators of business expansions and contractions, NBER Occasional Paper no. 103.

Neftci S.N. (1982), Optimal prediction of cyclical downturns, Journal of Economic Dynamics and Control 4: 225-241.

Qi M. (2001), Predicting US recessions with leading indicators via neural network models, International Journal of Forecasting 17.

Shepard N. (1994), Partial non-Gaussian state space, Biometrika 81: 115-131.

Sims C.A. (1989), Comment on Stock and Watson (1989), NBER Macroeconomics Annual: 394-39.

Stock J.H., Watson M.W. (1989), New indexes of coincident and leading economic indicators, in: Blanchard, O., and S. Fischer, eds., NBER Macroeconomics Annual, MIT Press (Cambridge, MA): 351-394.

Stock J.H., Watson M.W. (1991), A probability model of the coincident indicators, in Lahiri, K., and G. H. Moore, eds., Leading Economic Indicators: New approaches and forecasting records, Cambridge University Press (Cambridge, UK).

Stock J.H., Watson M.W. (1992), A procedure for predicting recessions with leading indicators: econometric issues and recent experience, NBER Working Paper Series, no. 4014.

Stock J.H., Watson M.W. (1999), A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, in: Cointegration, Causality, and Forecasting - Festschrift in Honour of Clive W. J. Granger, edited by R. Engle and H. White.

Stock J.H., Watson M.W.(2002a), Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business and Economic Statistics, 20, 147-62.

Stock J.H., Watson M.W. (2002b), Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association, 97, 1167–1179.

Timmermann A.G. (2005), Forecast Combination, in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting.
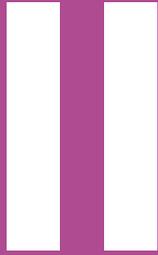
II | Statistical and Modelling Prerequisites

4 Data Availability, Frequency and Adjustment Techniques

Authors

Jan van den Brakel
Statistics Netherlands (CBS)/Maastricht University
j.vandenbrakel@mastrichtuniversity.nl

Sabine Krieg
Statistics Netherlands (CBS)
s.krieg@cbs.nl

Pim Ouwehand
Statistics Netherlands (CBS)
p.ouwehand@cbs.nl

Frank van de Pol
Statistics Netherlands (CBS)
f.van.de.pol@home.nl

Floris van Ruth
Statistics Netherlands (CBS)
frth@cbs.nl

Marc Smeets
Statistics Netherlands (CBS)
mset@cbs.nl

Piet Verbiest
Statistics Netherlands (CBS)
pvrt@cbs.nl

Table of Contents

4.1 Overview

It is the role of official statistics to produce undisputed figures, being a benchmark and safe haven in an ever extending sea of data. Constructing long and yet consistent time series is part of this task, as well as composing indicators from varying sources.

The main determinants of the quality of long series of composite indicators are

- Stability of research designs,

- Timeliness and proper definitions of data from several sources or regions,

- Equal frequency of data sources,

- Adequate derivation of indices and

- Wise decisions on outliers or turning points in seasonal adjustment.

In this chapter we focus on each of them. Firstly, components may not be stable enough due to changing classifications or some sort of a redesign. The problem can be mended with methods that are mentioned in Section 4.2. We give special attention to models for interrupted time series. An introduction to structural time series modelling is in an annex.

Another cluster of problems concerns incomplete or delayed data. Official statistics cannot always collect data according to the desired definition. Sometimes alternative data are freely available that can be used as an imperfect substitute or proxy. Two examples of alternative data are given in Section 4.3. We also pay attention to modelling data that come in too late or that are missing for some regions or branches of a classification.

Section 4.4 is about combining indicators of mixed frequency. We point out briefly some possibilities to do so and furthermore refer to the abundant literature.

Indices pose specific problems. An important issue is the detrimental effect of collecting data for only the first part of the month in order to speed up publication. In Section 4.5 we also direct attention to pitfalls with chain estimators and to the debate on how to cope with changing weights of index components, that is with a chain-linked or fixed base method.

Finally, sharp turns in a time series pose problems to seasonal adjustment. When should an outlier be considered as the beginning of a sharp turn? In Section 4.6 we plead for a conservative approach.

In this chapter we will often refer to model based approaches. However, it should be kept in mind that a good presentation of the information is vital. Modern animated graphics offer special possibilities. An example is the business cycle tracer (Van Ruth et al. (2005)) which shows several indicators simultaneously on their way through time. Such a business cycle clock is now being presented by several statistical agencies, among which Eurostat, the OECD and the statistical offices of Denmark, Germany, Korea and the Netherlands.

4.2 Quantifying discontinuities in sample surveys

4.2.1 Introduction

In order to forecast economic cycles, time series of macro-economic indicators should be as long as possible. Macro-economic indicators are typically based on repeatedly conducted sample surveys conducted by national statistical institutes. One important quality aspect of figures published with repeatedly conducted

sample surveys is comparability over time. To produce consistent series, national statistical institutes generally keep their survey processes unchanged as long as possible. It remains inevitable, however, to redesign survey processes from time to time to improve the quality or the efficiency of the underlying survey process.

Firstly, question wording of course affects the outcomes. Also response categories, the way in which these are presented (with or without interviewer, lay-out of the questionnaire, paper or web), the sequence of questions, previous questions, introductory text, explaining text, the length of the questionnaire; they all have effects on the outcomes, on the amount and direction of bias. Factual questions are less sensitive for these effects. The less latitude the respondent has for interpretation, the better. For enterprise surveys, these problems concentrate on internationally standardised nomenclatures, like the activity codes of enterprises (ISIC, NACE) and systems for international trade in goods (HS, CN, SITC), transport codes (NST) and classifications of services (MSITS: UNSD (2002, 2010)). See Eurostat (2009). Changes therein contribute to and constitute changes in measurement errors and bias.

Secondly, selective response is influencing the outcomes. Also, if a better, more complete, sampling frame becomes available, for an investments survey for instance, then estimates of investments will increase suddenly and the new figures are no longer comparable with the old time series.

Thirdly, alterations in data processing are influential. The introduction of a new imputation method, or an efficient data editing method like top-down editing or macro editing, all are potential causes of irregularities in the time series. For a classification of sources of errors in registers, we refer to Zhang (2012).

A redesign of a survey process generally affects the different sources of measurement bias in a survey and therefore has a systematic effect on the outcomes of the survey. These kinds of effects are referred to as discontinuities. If the method changes in either of the aspects mentioned above,

  (i) data collection instrument,

  (ii) target population or response percentage,

  (iii) data processing,

it should be checked whether the time series of target variables have been interrupted.

With respect to the measures we can take, we have several options, some of which are not applicable in specific situations. The best option is to apply parallel data collection following both designs, the old and the new. This may be done with a very important figure, like unemployment, for instance after the introduction of a web questionnaire.

A less expensive option is available if only classification variables are affected. In the case of the introduction of a new version of the NACE for instance, we may aim at calculating outcomes with both versions of the classification variable. This implies that we need a questionnaire that supports both classifications in the transition period. Moreover, for a good comparison, sample size in publication cells should be large enough for both the old and the new classification. Therefore, in some cells a larger sample size will be needed in the transition period.

A less expensive alternative for a parallel data collection or parallel recalculation is to "glue" the series before and after the interruption with a step parameter in a time series model. In order to estimate the step parameter one needs an accurate estimate for the first time point next to the interruption and for that a sufficiently long time series is needed. If available, an auxiliary time series that is strongly correlated with the target series may help to bridge the gap more accurately. This approach will be treated in detail in the following subsections.

Lastly, we have of course the zero-option to do nothing. For less important target variables, a statistical office may decide to communicate that the time series has been interrupted and a new series has been started, because specific methodological changes or improvements were necessary. In such a case, the new time series unfortunately may not be comparable with the old series on the same target variable.

However, in an ideal survey transition process, the systematic effects of the redesign are explained and

quantified to avoid confounding real developments of the indicators of interest with discontinuities induced by the redesign of the survey process. This keeps series consistent and preserves comparability of the outcomes over time. To sum up, one or more of the following four options may be available,

a) parallel data collection with the old and the new design,

b) parallel data processing in the old and the new way,

c) application of an interrupted time series model,

d) do nothing and take the blame for a discontinuity in the time series.

In survey methodology, time series models are frequently applied to develop estimates for periodic surveys. Blight and Scott (1973) and Scott and Smith (1974) proposed to regard the unknown population parameters as a realization of a stochastic process that can be described with a time series model. This introduces relationships between the estimated population parameters at different time points in the case of non-overlapping as well as overlapping samples. The explicit modelling of this relationship between these survey estimates with a time series model can be used to combine sample information observed in the past to improve the precision of estimates obtained with periodic surveys. This approach is frequently applied in the context of small area estimation, see e.g. Rao and Yu (1994), Pfeffermann and Burck (1990), Pfeffermann and Bleuer (1993), Pfeffermann et al. (1998), and Pfeffermann and Tiller (2006). Time series models are also appropriate to account for differences in measurement bias in rotating panels (Pfeffermann (1991), Van den Brakel and Krieg (2009)) and for discontinuities induced by a survey redesign, (Van den Brakel et al. (2008) and Van den Brakel and Roels (2010)). Some other key references to authors that applied the time series approach to repeated survey data to improve the efficiency of survey estimates are Scott et al. (1977), Tam (1987), Binder and Dick (1989), Binder and Dick (1990), Bell and Hillmer (1990), Tiller (1992), Harvey and Chung (2000), Feder (2001), and Lind (2005).

Before structural time series models emerged, a regression approach to time series has been common practice Ostrom (1978), and it still is widely used. This approach has developed throughout the years into RegArima models, coping with autocorrelation by taking first order differences among other things. Intervention parameters can be estimated also (Box and Tiao (1975)). Nowadays this RegArima approach is often used as a first step, including estimation of discontinuities, before seasonal adjustment filters are applied (Bureau of the Census (2013); Caporello and Maravall (2004); Grudkowska (2013)). However, for modelling discontinuities we have a preference for structural time series models because they offer both good interpretability of parameters, like in seasonal adjustment heuristics, as well as statistical models that can be tested.

The need as well as the various possibilities to quantify discontinuities, is discussed in subsection 4.2.2. In an annex to this chapter, a general framework for modelling discontinuities in time series observed through repeatedly conducted surveys is constructed. There we elaborate on the state-space representation of these models. Also the Kalman filter, commonly used to analyse state-space models, is briefly reviewed. It is also indicated how external quantitative information about discontinuities can be used in the model. Appropriate software to estimate state-space models is briefly reviewed.

4.2.2 Quantifying discontinuities in sample surveys

Changes in a survey process can have large systematic effects on the outcomes of a survey. An experiment with a new questionnaire for the Dutch Labour Force Survey (LFS) in 2000 showed that the unemployed labour force, defined as the ratio of the total unemployed labour force to the total labour force, dropped from 3.1 to 2.4 percentage points (Van den Brakel and van Berkel (2002)). An experiment with alternative data collection procedures in the LFS, conducted in 2001, showed that the estimated unemployed labour force under computer assisted telephone interviewing is 1.1 percent points smaller compared to computer assisted personal interviewing, Van den Brakel (2008). Finally an experiment conducted to estimate discontinuities due to a major redesign of the LFS in 2010 showed that a change-over from computer assisted personal interviewing

to a mix of computer assisted telephone and personal interviewing in combination with the introduction of a new questionnaire increases the estimated total unemployed labour force from 420.000 to 475.000 people, Van den Brakel and Krieg (2012). These three examples illustrate that modifications in a survey process can have substantial and significant effects on the figures of a survey. It is therefore crucial that discontinuities due to modifications of a survey process are quantified and that data users account for these effects in their analyses.

There are various possibilities to quantify the effect of a survey redesign, see Van den Brakel et al. (2008)) for an overview. The required methods depend on the phase of the survey process that is changed. In cases where the underlying sample data remain the same, the differences can be investigated by recalculation, for example the introduction of new editing, imputation or estimation methods. Also a new economic activity classification system in business surveys will generally result in discontinuities in time series. An example is the change-over from the NACE Rev 1.1 to the NACE Rev. 2 in 2008. The effect of such a new classification system can be quantified by adding the new classification system to all units in the sample frame or at least in the sample. As a result, a double coded sample or sampling frame is obtained for a particular reference period, and appropriate design-based domain estimators can be constructed under both classification systems. This enables the quantification of the effects of the change-over. See Van den Brakel (2010) for details and an application to the change-over from the NACE Rev 1.1 to the NACE Rev. 2.

A bottleneck in this approach may be that data editing has to be carried out twice too, requiring more statisticians than are actually available. Then a simplified procedure may be followed, involving a transformation of publication figures. Results according to the old NACE after data editing can be distributed over relevant cells of the new NACE. In order to do so an enterprise transition matrix from the old classification to the new classification has to be constructed that holds the proportions that stay and that move to specific cells.

When data collection procedures are affected, the new micro data are not consistent with the past. In these cases the effect of the change-over can be quantified by conducting a field experiment where the regular and new survey designs are run concurrently. Under a well-designed parallel run, two estimates for the variables of interest are obtained for the same reference period; one under the regular and one under the new approach. The contrast between both estimates is a direct estimate for the discontinuity induced by the redesign of the survey. See e.g. Van den Brakel (2008) for applications and details of designing and analysing this kind of large scaled field experiments. The old design may be maintained for several periods with a sample that is small, but just big enough to discern relevant changes in the level of unemployment. If explanations are to be given for the changes that occur, a larger sample will be needed in order to be split up in explanatory subsamples, like age groups.

A parallel run is not always tenable for a national statistical institute due to budget constraints. Moreover, doubling the data collection is not possible with enterprise surveys, as the largest enterprises are included in the sample with probability one. It is not feasible to ask these enterprises to fill out two different questionnaires on the same subject. In such cases a time series modelling approach can be considered as an alternative. In this case the evolution of the series of the variable of interest is modelled with an appropriate time series model. To quantify the effect of the redesign, an intervention variable that describes the moment of the change-over from the old to the new design is added to this model. Under the assumption that the components of the time series model, other than the intervention component, describes the evolution of the series reasonable well, the regression coefficient of the intervention variable will measure the systematic effect induced by the redesign of the survey. In an annex to this chapter, we will give an introduction to the use of a state-space framework for modelling time series observed with repeatedly conducted survey samples. Intervention models are developed to construct uninterrupted series for the target variables. It is discussed how additional information about discontinuities, for example from a parallel run, can be included as prior information in these models.

4.2.3 Adjusted series, back-casting

In the annex, a state-space framework to construct consistent series is proposed. In some cases, however, a simple regression approach may also suffice (Ostrom (1978)). After having analyzed the intervention model, the observed series can be adjusted or corrected with the estimated discontinuities, in a way comparable to constructing seasonally adjusted data. In the case of a level intervention, for example, the time series after the moment of the survey transition can be adjusted for the estimated discontinuities with $\tilde{y}_{t,k} = \hat{y}_{t,k} - \hat{\beta}_k$, with $\tilde{y}_{t,k}$ the new series and $\hat{\beta}_k$ the amount in which the new series exceeds the old one in category or stratum k. Adjusting the new time series cannot be continued for a very long time, of course. Therefore we prefer the alternative, that is to adjust the series before the survey transition, with $\tilde{y}_{t,k} = \hat{y}_{t,k} + \hat{\beta}_k$. This may be called back-casting. As stated before, an auxiliary series can be added to the model as a proxy, in order to get a more accurate estimate of the intervention step. When for instance a redesign has interrupted the series from a structural business survey, VAT figures may be incorporated in the structural time series model to help bridge the gap.

However, adjusting the old series and favoring the new series implies that we have to argue why the new figures are better, which may be hard when redesigns are motivated by cutting costs. Therefore it is important that official statistics maintain or improve quality when redesigning surveys. Quality can be measured partially as representativety of the responding part of the sample (Schouten et al. (2011)).

As an alternative for a correction with $\hat{\beta}_k$, filtered or smoothed estimates for the target parameters can be used. Intervention models explicitly account for discontinuities. Filtered estimates for the target variables are therefore not affected by the systematic effect of the change-over.

The two aforementioned approaches are useful if consistent series are required to produce uninterrupted series of indices or if consistent macro indicators are required as input for other models used to analyze economic cycles.

If the state-space framework itself is used to analyze economic cycles, then the intervention components developed in this chapter can be included in the model to avoid model-misspecification and distortion of the estimated economic cycles by the discontinuities in the input series.

For back-casting in the context of RegArima models we refer to Sartore and Caporin (2006). These authors focus among other things on aggregated time series for the EU. When one of the EU-countries revises its series, this invokes recalculation of the back-casting exercise for the aggregated EU series.

4.3 Missing, delayed or incomplete information: What alternatives do we have?

Composite indicators cannot be built without some sort of statistical information. However, we don't always have the right information, that can straightforwardly be added up or raised to population figures. The Organisation for Economic Co-operation and Development (OECD (2008)) gives a good exposition about various imputation methods as the solution for missing data problems. In addition, we want to highlight situations where imputation is less relevant. If there is no direct information on your target variable, maybe proxy information is available which is known to be highly correlated to the figure that we want to estimate. Secondly, if the desired information becomes available after a long time, that is if it comes too late, auxiliary variables that become available sooner may be used for a provisional estimate of the current situation. Finally, sample size of a survey may be too small for a detailed breakdown. Information on a specific region or stratum may even be completely missing. In that case a highly correlated variable from a register, or data from similar strata or areas, may be helpful to obtain more detail. If there are no data on a specific area and no better option is

available, even a completely synthetic estimate could be made. We consider three cases to illustrate different solutions,

1. newly realised buildings,

2. consumer confidence and

3. the order book of establishments. The section ends by reflecting on administrative data versus surveys.

4.3.1 Newly realised buildings

The best option to monitor the building production is to maintain a register of addresses, of course. Such a register should be kept up to date by municipalities or any other type of institution that earns money from lending services to these addresses, like maintaining roads and other infrastructure, providing energy, collecting garbage, etc. Nations can simplify administrative tasks for such enterprises by offering one unifying generic register of addresses.

The ideal to have a timely and complete address register is difficult to reach, though. If an address register is not available, a survey may be set up, asking local municipalities or counties to assess the current stock of houses and the proportion that was built after the last occasion the survey was held, for instance ten years ago.

These figures can be updated with a yearly survey on building permits that local municipalities have granted. With these updates, the fact should be taken into account that not every permit will result in a completed building. In some highly regulated countries, people have the right to object against building permits for a specific period of time. To counteract this risk, sometimes more than one version of a building permit is delivered. Moreover, adverse economic conditions may inhibit and delay building plans. Hence an estimate is needed of the fraction of the building permits that will be realised within the first, second and later years. Also the limited scope of a survey on building permits should be taken into account. Usually a lower threshold of for instance €50.000 is maintained and temporary buildings are ignored.

In addition to the building permits survey one may monitor the amount of concrete that is being used in order to estimate an increase or decrease in the percentage of building permits that results in a completed building. Also the turnover of building enterprises, and short term changes therein, known from VAT data, can be helpful to assess this parameter. We will discuss administrative data in general at the end of this section.

4.3.2 Consumer confidence and social media data

Because consumer confidence is an early indicator for business tendency, several countries carry out a survey on this subject. Daas and Puts (2014) recently found that the time series of consumer confidence is strongly correlated with the sentiment in part of the social media, especially Facebook and to a lesser extent Twitter. Consumer confidence seems to be leading and social media sentiment lags a bit behind. These findings seem promising, but more research has to point out whether consumer confidence can be derived from social media on a regular basis, for instance weekly, in the future. If so, only a yearly survey would be needed to corroborate and tweak the estimate.

This is just an example of the potential of "big data". For instance Google Trends and Google Correlate offer a toolbox for trend watching and exploring the internet for statistical purposes. The underlying methods are also referred to as "data science".

4.3.3 New orders for establishments

Another useful indicator for economic prospects is the value of the order book of enterprises. "Industrial new orders are among the leading series used for the widely monitored OECD Composite Leading Indicator", state De Bondt et al. (2013). Because this information is no longer an obligatory part of the Eurostat short term statistics (STS), the European Central Bank (ECB) continues the series, notwithstanding the fact that some countries have stopped collecting these "hard" data on new orders, due to budget cuts and pressure to reduce the administrative burden.

An interesting question is whether data from countries where the hard information is still available, can be used to obtain valid estimates for the countries where these data are lacking. This problem can be considered as a special case of small area estimation (SAE). In small area estimation (Eurostat (2013)) information on an intermediate level, like region or branch of industry, is used to reinforce survey estimates. Especially in the areas where few survey observations are available, the model based estimate will get a large weight compared to the survey estimate. In the extreme case where no information is available one may rely completely on the model based prediction, which will then be called a synthetic estimate. Most literature on SAE is on data where every observation is equally important, as with persons and households. A reference specific for business surveys, where the data are not normally distributed, but skewed, is Krieg et al. (2012).

Instead of applying the rather involved SAE methods, or in a preparatory phase before doing so, one may make common sense estimates for countries or regions where hard information on the order book is missing. This may be done for instance by selecting comparable areas, taking a stratification by branch of industry into account. Also other early indicators for the tendency of the market could be taken into account, such as vacancies and investments.

The ECB has developed illuminating time series models that cope with partially missing information about investments. They do not only rely on hard information on new orders, which lacks for some countries, but also use information that is available for all countries, although maybe a bit less leading. Next to the leading survey indicator "managers' assessment of the current level of orders books", also the hard "Industrial turnover index in manufacturing" is used as a second indicator. Several transformations, such as lags and first order differences, have been tested and incorporated into the model where useful. De Bondt et al. (2013) also report various robustness checks.

4.3.4 Administrative data or surveys?

In order to reduce the administrative burden, data collecting agencies are put under pressure to replace surveys with the use of administrative data as much as possible. This means that instead of survey variables that can be defined in any appropriate way, administrative data are obtained, which have to be accepted as they are. Restrictive is also that registers generally offer a small number of variables which are relevant for statistics, whereas the number of variables in a survey can be larger.

An advantage of administrative data is that registers offer many, sometimes almost all, members of a population, whereas surveys, due to cost restrictions, concern only a limited sample. Hence with surveys weighing and minimising sampling inaccuracy are important issues, whereas with administrative data selectivity of the missing part of the population is an important issue.

Various mixed designs emerged, bringing along estimation issues. For instance one may rely on VAT data for smaller enterprises, whereas a survey with a longer questionnaire is used for the larger enterprises. Another hybrid design emerges when a yearly survey sets the levels and an administrative source, like VAT, is used for short term extrapolation. However, some administrative sources are too slow to be used for a short term indicator. Therefore some countries use a survey to catch the short term general trend.

Not only survey data, but also administrative data can hold erroneous information (Zhang (2012)). Register holders will have a keen eye for their core variables, like value added in a tax setting or the number of employ-

ees in a social security setting. However, register holders will be less strict when it comes to for instance the timing of transactions or the categorisation of employees as temporary or permanent. Needless to remind the reader that survey respondents make errors too.

In sum, switching to administrative data, in part or completely, poses many different additional problems compared to surveys.

4.4 Mixed frequency data and delayed results

Policy makers rely on several indicators for the business cycle, but these indicators will not become available simultaneously. Indicators that appear monthly can be available earlier than quarterly indicators. Moreover, some monthly indicators require a longer fieldwork period or more processing time than others. This ragged edge of the available time series, some of them already available in the most recent month, and others not, is a problem that has been tackled by time series modeling in various ways recently.

Quarterly data can be transformed to monthly data in various ways. The time series analysis program Eviews has a tutorial for frequency conversion that shows several options, one of them being linear interpolation. Other methods of temporal disaggregation are Chow-Lin interpolation (Chow and Lin (1971)) and the Denton method (Denton (1971)). Next to temporal disaggregation, one can also overcome the multi-frequency problem by aggregation to the slowest, that is yearly, frequency, but this does not provide very recent now-casts of the present situation, because yearly figures become available quite late.

In order to take advantage of the latest information in the "ragged edges", bridge models are often used for now-casting (Foroni and Marcellino (2013)). In this approach the high frequency, usually monthly, data are used to create forecasts or a now-cast for the current month. The equations for forecasting may then be used to extend quarterly time series with estimates for the latest months. Durbin and Quenneville (1997) proposed a state-space framework to benchmark time series observed on a monthly frequency to more reliable series observed on a quarterly or annual frequency.

A more recent development is the use of distributed lag polynomials for mixed data sampling (MIDAS). Alternatively, the state space models that were described at the beginning of this chapter may be used, extended to mixed-frequency vector autoregression (VAR) and factor models. An advantage of the latter type of models is that estimates for various frequencies can be forced to be consistent. For details the reader is referred to Foroni and Marcellino (2013) and the literature and software mentioned therein.

Statistical agencies also rely on imputation of the missing, most recent, data, as a practical alternative for structural or reduced form statistical models. Statistical imputation can be performed with simple time series models or other ways to impute or extrapolate the missing data with instrumental variables. Whereas statistical models give more opportunity to compare and test several ways for predicting missing information, imputation procedures are easier to handle in the production of statistical data. This is in important advantage when figures should be produced as soon as possible. The U.S. Leading Economic Index (LEI) for instance, which has been published since 1968, suffered from "ragged edges" with some of the constituting time series. McGuckin et al. (2007) proposed an imputation procedure that makes better use of the real time financial data. As a result, the official "Conference Board" U.S. Leading Index can be published two weeks earlier than before. The Conference Board implemented this procedure in its 2001 benchmark revision of the LEI. See also Conference Board handbook on Business Cycle Indicators (2001).

4.5 Incomplete data and indices

4.5.1 Introduction

In the ideal case we get a 100% response throughout the whole observation period, not during just a part of the observation period. Also overdue response from previous periods is to be ignored in reports about later periods. Practice, however, is different. Response rates may be low at the earliest publication occasions. Moreover, some institutions collect data from the middle of one month to the middle of the next month and label the results as representing the whole latter month. Other institutions erroneously count overdue response as representing the month when it was received, whereas it should be counted with an earlier month, the month that the respondent reported about. We pay for these inconveniences and errors with several potential biases. Developments of not responding units are not represented in the published figures and figures get published with a more recent tag than is justified by the contents. In the present section we will elaborate on these problems.

In the third subsection we will look into the problem of following changes in the population of enterprises when an index like turnover is to be computed.

Fourth, we note that the economy is changing rapidly. Frequently, new products like smart phones appear on the market while others, like tube televisions, disappear. In order to make "the best" estimates for price and volume indices, weighting schemes should be kept in line with the dynamics in the economy. To reach this in practice, often the previous year is used as base year in index formulae. A consequence of applying such a "moving base year" is that a consistent time series of indices cannot be made. A solution for this is the use of chain linked indices.

4.5.2 An observation period that does not match with the reporting period

For some time now, an important issue in European statistics has been the improvement of the timeliness of statistics. This research is driven by the demands of the ECB, Eurostat, financial institutions and economic analysts for the faster publication of data. One method for achieving this goal is to take measurements only during a part of the period on which one will report. For example, only during the first two weeks of a month. The data for the observed part are assumed to represent the whole period. Thus earlier publication is achieved. The problem which then arises is how the accuracy of a statistic measured by this method is affected. All developments in the ignored part are missed. Surprisingly, very little research into this problem has been performed. This is all the more relevant as a large number of statistics is routinely produced in this manner. Lack of data probably has been a major cause of this deficit in research. We will summarise here results of a study into the influence of timing and length of the measurement period on monthly price indices by Van Ruth (2002). Specifically, it has been studied whether there are weeks in the month which yield more accurate monthly indices than other weeks. Further, the influence on the accuracy of lengthening the price collection period is examined as well.

New data collection techniques have resulted in data becoming available at higher frequencies, for example prices. These are used to construct inflation measures such as consumer price indices (CPI). Virtually every country in Europe publishes a monthly consumer price index. The measurement period varies widely, from one day in the middle of the month to almost the whole of the month. In the Netherlands some price data on a daily or weekly basis have become available. These have been used to investigate the influence of using different measurement periods on the price index series.

Three types of data are available: petrol prices, traditionally collected prices for items that exhibit frequent price variations, such as vegetables, and so-called scanner data from super-market chains. From each of

these series an index based on the whole month and on different parts of the month has been constructed. These then were compared. It will be attempted to generalise the results to other statistics.

It is difficult to make a general judgement on the use of shorter measurement periods for the production of price indices. Indices based on measurements during just one week of a month result in large monthly deviations from the index as based on data for the whole of the month. Average errors are generally as large as or larger than average monthly changes in the reference series. On top of this, maximum single errors can be very large, up to 35%. Errors in the monthly year-on-year price change rate, usually the statistic of interest, are large as well. The average error for single week indices is about half the average monthly rate of the whole month series. Therefore, in that case, inflation rates will possess a large amount of uncertainty.

On the other hand, for conventionally measured price indices, the price development in time of the different indices is quite comparable. So if the long-term evolution of inflation is the main interest, measurement in one week can be acceptable. And using the first two weeks of a month will then result in improved accuracy. Also, developments missed in one month will usually show up a little later. This is not so for price indices based on scanner data. The inclusion of special offers, usually present in single weeks, means that important price changes can be either missed or overstated in single week indices. This is reflected as well in the relatively very large error in the single week inflation rates. The average error is about as large as the average monthly inflation rate of the reference series. It is clear that, for the products considered here, short price collection periods cause a large amount of relevant information to be missed. Therefore, if the short-term development of prices is of primary interest, collecting prices in only a (minor) part of the month will probably be unacceptable. This is especially so if stringent accuracy requirements are present.

All these problems are mitigated when longer period averages are used. Two-week averages more or less halve the errors considered, and three-week averages bring these down even more. This last method also diminishes the influence of price variability on the magnitude of the errors. It has been shown that, as was to be expected, both average monthly errors and maximum errors increase with increasing price variability. The effect is strongest on the level of the maximum errors as measured in the individual item's time series. For the classes of items with the lowest price variability, there is actually not much difference in performance between the different measurement periods. Unfortunately, these constitute only a minority of the sample.

The acceleration of price index measurement by considering only part of the reference period clearly carries a cost in accuracy. How serious this cost is, and if the improvements in timeliness are worth it, depends on the users of the statistics. Economists from the ECB, an important user, have stated that frequent revisions of the HICP of more than 0,1 percentage point would be as harmful as late publication . If for example petrol prices are considered, this already proves to be a difficult constraint. In the Netherlands, petrol has a weight of 2.7% in the CPI and even in the best case an average error of 1,4 in the index is present. The average error in the CPI resulting from this would be about 0,04 percentage point. Thus just one of the items, with a relatively small weight in the overall index, already consumes a large part of the margin of error.

Finally, it is interesting to look at the concept of shorter measurement periods in a broader perspective. The question is, whether given these results for price indices, the method is applicable on other statistics as well. It depends on the type of statistic. For statistics which do not fluctuate very much over shorter periods and for which the general development in time is the most important, including only part of the reference period seems to be warranted. These conclusions are in line with the ILO-CPI manual on price indices as outlined in International Labor Organization (2003)).

4.5.3 On chain estimators and changes in population or sample

Establishment surveys should represent the population as closely as possible. Therefore the sampling scheme should not only reflect the fact that some enterprises disappear, but should also allow the inflow of new enterprises and enterprises that have moved to the activities that are the subject of the present survey. However, this may result in serious distortions of an index that is derived from that survey. To illustrate this we discuss

an economic indicator like turnover, that is represented as a percentage of a base year, which is set to 100. In such a case the index for period t, I^t, should be

$$I^t = 100\frac{Y^t}{Y^0},$$

with Y^t the aggregate at period t, for instance turn over, and Y^0 the same figure in the base period. This index may be thought of as a product of growth rates, $G^{t-1,t} = Y^t/Y^{t-1}$, for all intermediate periods,

$$I^t = 100G^{t-1,t}G^{t-2,t-1}\ldots G^{0,1}$$
$$= 100\frac{Y^t}{Y^{t-1}}\frac{Y^{t-1}}{Y^{t-2}}\ldots\frac{Y^1}{Y^0}$$
$$= 100\frac{Y^t}{Y^0}.$$

Rewritten in this way, it is readily seen that this product of growth rates is equivalent to the ratio of the current aggregate to the aggregate at its base period.

Things may go wrong when the population changes. Suppose that a growth rate can only be computed with that subsample of the population for which comparable data are available for both subsequent periods. Then the current estimate for the previous occasion, $Y^{t-1,t}$, is not exactly equal to the estimate $\tilde{Y}^{t-1}$ at the previous occasion. Therefore the desired index is polluted with a bias factor that is not automatically corrected,

$$I^t = 100G^{t-1,t}G^{t-2,t-1}\ldots G^{0,1}$$
$$\approx 100\frac{Y^t}{Y^{t-1,t}}\frac{\tilde{Y}^{t-1}}{Y^{t-2,t-1}}\ldots\frac{\tilde{Y}^1}{Y^{0,1}}$$
$$= 100\frac{Y^t}{Y^0}\left[\frac{\tilde{Y}^{t-1}}{Y^{t-1,t}}\frac{\tilde{Y}^{t-2}}{Y^{t-2,t-1}}\ldots\frac{\tilde{Y}^0}{Y^{0,1}}\right]$$
$$= \text{index}(t) * \text{biasfactor}(t).$$

The same problem arises in any situation where an estimate of Y^t takes another value when compared with Y^{t-1} than when compared with Y^{t+1}. This can be the case if an error is edited in between the publication of $G^{t-1,t}$ and $G^{t,t+1}$. Also robust estimation methods can introduce a small bias, usually an underestimate of the growth rate, that will build up to become a large bias as the chain of growth rates gets longer. Van Delden (2006) studied and quantified the various components that contributed to bias in the index of a specific NACE code, the supermarkets.

These bias inducing problems with the present chain index become visible as soon as a new basis for the index is chosen and the current Y^t appears to be inconsistent with Y^0I^t. They can also become apparent before that time, when subseries are inconsistent with the generic index for both subseries. Suppose that next to the economic growth rate two subseries are published, namely the domestic and foreign growth rate. Suppose we have the following gross turnover for periods 0, 1 and 3.

Table 4.1: Turnover and index values

period	Turnover and			index values		
	domestic	abroad	total	domestic	abroad	total
0	80	20	100	100	100	100
1	80	5	85	100	25	85
2	60	5	65	75	25	65

Then the index values are as the right hand side of table 4.1 shows. However, activity codes for several enterprises have changed. Therefore, the following corrections were carried through. Please note that the corrections for domestic and foreign turnover add up to total turnover, i.e. they are consistent.

Now growth rates are found that differ from those that appear from table 4.1. The domestic growth rate from period 1 to 2 is no longer 0.75, but (60-10) divided by (80-2), that is 0,64. (Enterprises contributing to table

Table 4.2: No longer valid turnover (to be ignored in comparison with the next occasion)

period	domestic	abroad	total
0 to 1	0	5	5
1 to 2	2	3	5

Table 4.3: Newly added turnover (to be ignored in comparison with the previous occasion)

period	domestic	abroad	total
0 to 1	0	0	0
1 to 2	10	1	11

4.2 have left the present activity code and should not be considered in the comparison with the first period where their observations are no longer available in the present NACE section. Enterprises in table 4.3 just entered the present NACE code and should not contribute to the comparison with the previous period, when they were classified elsewhere.)

Table 4.4: Inconsistent index values after data mutations

period	domestic	abroad	total
0	100	100	100
1	100	33	89
2	64	67	60

After computing growth rates and deriving index values from that, it appears that the total index value at period 2 is no longer consistent with the index values of the constituting categories: 60 is not in between 64 and 67, though it should be if the figures were to be consistent.

Such inconsistencies cannot always be avoided, but some choices make their occurrence less influential. Firstly both estimates in a growth rate, Y^t and Y^{t-1}, should be consistent with respect to the treatment of errors. Secondly, it is advisable to benchmark the series at regular intervals. In the case of the turn-over index one can use quarterly or yearly VAT data for this purpose. (In some countries VAT data become available too late for short term indicators, at least for small enterprises.) If benchmarking is not feasible, one should avoid those (robust) estimation methods that tend to reduce the influence of positive outliers only. Moreover, moving enterprises from one activity code to another should be avoided or postponed to the next base year (at the cost of less validity). If a benchmark is available, activity codes can be adapted more often, though.

The issue of actuality versus consistency does not only pop up at the level of sample selection criteria, but also plays a role in determining weights in price indices, as will be discussed in the next subsection.

4.5.4 Chain linked versus fixed base price indicators

Changes in prices and volumes are important indicators for judging the performance of an economy. The consumer price index and the real growth of gross domestic product are well known examples. For this purpose national accounts of an increasing number of countries make use of a chain linked method in which the weights for prices and volumes of the components of the economy are updated annually. Generally Laspeyres type indices are used for price and volume estimates . In a Laspeyres index formula individual price and volume changes are weighted with the values of the concerning transaction in a "base year". Generally speaking there is a choice between a fixed base year and a moving base year.

With the method of fixed weights for a series of years, the weights are derived from a single year in the past. An advantage of this method is that in longer series of level estimates of deflated components of aggregates at prices of the base year, these level estimates exactly add up to the deflated aggregate. However,

a very serious disadvantage is that volume changes of aggregates are calculated with outdated weights. This disadvantage is especially severe when relative prices change rapidly. As a result, economic growth can be significantly overestimated. Also appearance of new products (mobile telephony, tablets) or the disappearance of others (tube television) can lead to distortion of estimates of economic growth.

Applying a moving base year means that weights are updated every year and are usually derived from the previous year. Since those weights are more up-to-date, a better approximation of the "real world" price and volume changes is obtained than with the method of fixed weights. However due to the changing weights, compilation of time series of indices is not straightforward. Time series results can be obtained by chaining of indices, i.e. multiplying separately estimated year-to-year price and volume indices. (The time series starts in a specific year for which the index is set to 100.) An important advantage of the chain linking method with moving weights is that the above mentioned overestimation of growth rates is avoided. There is also a disadvantage with the time series of indices when a moving base year for prices is applied. In the time series in terms of prices of a specific reference year, the deflated components of an aggregate no longer exactly add up to the deflated aggregate. As a result "mathematical discrepancies" will appear that cannot be removed without distorting the underlying "actual" volume and price movements.

More details, tips and tricks can be found in Brueton (1999) and in leading manuals from the International Labor Organization (2003), Eurostat (2001), Eurostat (2009; 2015) and the International Monetary Fund (2010).

4.6 Seasonal adjustment in times of strong economic changes

4.6.1 The problem and the options

Starting in 2008 the world has been subject to a severe economic crisis. A collapsing confidence in the solidity of sub prime mortgages hit the financial sector, and later spread to most parts of the real economy. In most countries, the strongest fluctuations took place in the period 2008-2010.

Due to these fluctuations, the economic crisis had its implications for the seasonal adjustment process. On the one hand, seasonally adjusted figures help in interpreting short term economic developments. On the other hand, the seasonal adjustment process itself is subject to this increased volatility in the data. Producing accurate seasonally adjusted figures is more difficult under these circumstances.

Specifically, there are two problems making seasonal adjustment more difficult:

- Due to the strong fluctuations and rapid changes estimating the seasonal pattern is more difficult, especially at the end of a series. At the same time, these are often the most important figures of a series. The main question is whether these movements at the end of a series are temporary or whether they imply more structural economic changes.

- The seasonal pattern itself may change due to the crisis as well. Again, this can be a temporary change or a more permanent one. At the time seasonal adjustment is carried out, it would be good to know whether the change in the seasonal pattern is a temporary or more permanent change, or whether we are dealing with just heavy fluctuations in the data. The problem is that we can not know at that point.

Now that the crisis has been going on for quite a while, it is important to evaluate the approaches applied to deal with the above problems. In the following subsection we describe seasonal adjustment practice during the crisis, and summarize lessons learned and give recommendations for future steps.

In the following we first give an overview of the series used for this research. Then we describe how the crisis first became visible in the series and could be recognized. Next the influence of the crisis on the sea-

sonal adjustment process is described. Finally we compare approaches of various countries, draw tentative conclusions and give recommendations.

Several software packages are available for seasonal adjustment. In the context of state space models and time series we mention Stamp, SsfPack and Ox (Koopman et al. (2009); Koopman et al. (2008)). For seasonal adjustment often a RegArima approach is applied first, to be followed by a filtering method for seasonal adjustment (Findley et al. (1998)). This can be accomplished with X13-Arima-Seats (Bureau of the Census (2013)) or Tramo-Seats (Caporello and Maravall (2004)). Both approaches are implemented in Demetra+ (Grudkowska (2013)). For a general treatment of seasonal adjustment using X-12-Arima, we refer to Van Velzen et al. (2011). Eurostat (2009; 2015) gives general guidelines for seasonal adjustment.

4.6.2 The case of the economic crisis of 2008-2010 as reported in the Netherlands

Overview of series

We investigated a wide range of series that are seasonally adjusted at Statistics Netherlands. The following gives an overview of the series we looked into:

- National accounts (quarterly, e.g., GDP, imports, exports)

- Macro economy (monthly, e.g., industrial production index)

- Confidence indicators (monthly, e.g., Consumer/producer confidence index)

- Labour market (monthly, e.g., vacancies, unemployment benefits)

Presence and identification of the crisis

The first indication of what later turned out to be a crisis came with the consumer confidence index of September 2007. A sharp decline was witnessed at this moment, which can be seen in Figure 4.1.

After that, the first signs of a crisis in the real economy were not seen until the third quarter of 2008. (See GDP series in Figure 4.2.)

For quarterly national accounts series, the start of the crisis was relatively clear. In the second quarter of 2008, several important indicators (GDP, imports, exports) showed a decline in seasonally adjusted figures. In the third quarter, in all other quarterly series a decline could be seen. Since the quarterly series are published only 45 days after the quarter has finished, this decline did not become apparent until October 2008.

This strong decline is illustrated in Figure 4.2 and Figure 4.3. The decline in the second quarter of 2008 can be clearly seen in Figure 4.2. While the previous quarters exhibited high quarter-on-quarter growth rates, the second quarter suddenly drops. At the same time the year-on-year growth rates (Figure 4.3) for the same quarter are still quite high (3,3%). However, the year-on-year growth is approximately equal to the sum of the four most recent quarter-on-quarter growth rates. Since the previous three quarters were strongly positive, the year-on-year growth rate only decreases only partly when adding the second quarter of 2008. The decrease is thus a turning point after a period of growth.

In 2005, the year-on-year growth rate shows a strong decline as well. Unlike in 2008, here the decline is only brief. In order to judge whether the economy was in a recession at that moment, we should look at the quarter-on-quarter growth rate. According to the definition, a recession requires two consecutive quarters with a negative quarter-on-quarter GDP growth rate. In 2005, however, there was no negative growth.

Figure 4.1: Consumer confidence index

Consumer Confidence Index
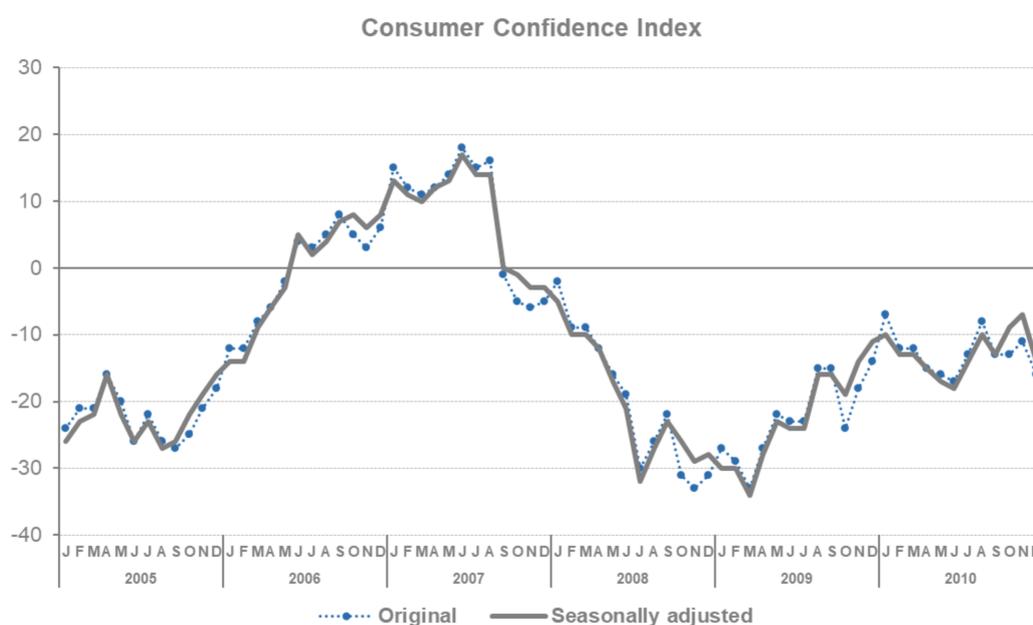
Source: Authors' own calculations

Figure 4.2: Quarter-on-quarter percent change in GDP

GDP: Value at 2000 prices
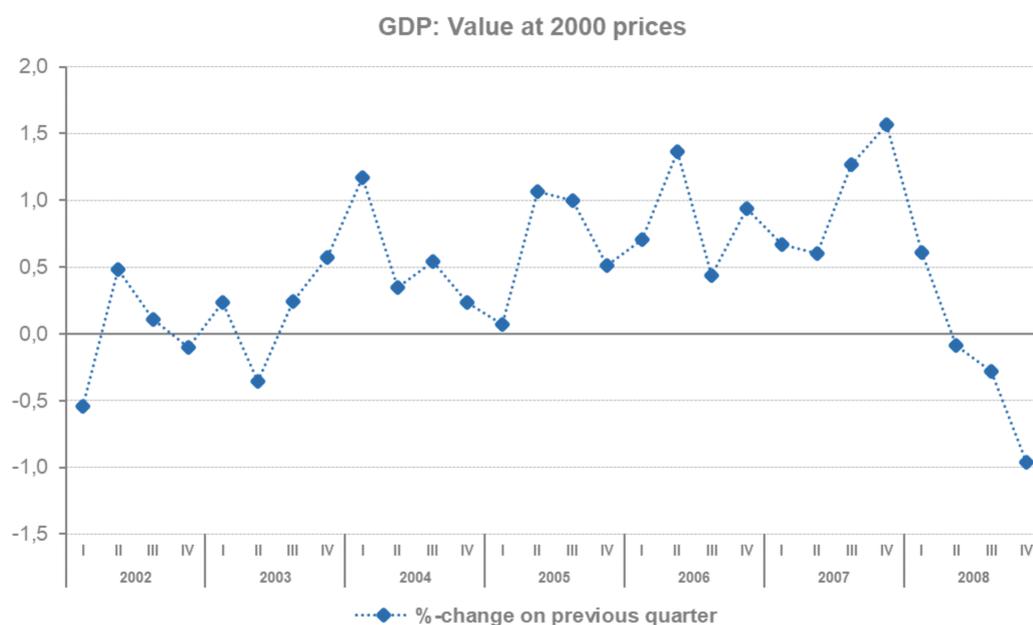
Source: Authors' own calculations

The judgement that the strong declines should be interpreted as a crisis was based on consensus of experts within and outside Statistics Netherlands (economists, academics). The above definition of a recession was applied. According to this definition, there was a recession in the third quarter of 2008 (See Figure 4.2).

For the monthly series, there was a clear influence of the crisis on the further development of these series. The regular seasonal pattern was clearly less visible than before the start of the crisis. The severity of the crisis caused the series to decline rapidly. Because of this, no seasonal peaks could be seen, but instead, deeper troughs. It is not clear whether actually the seasonal pattern was not present or whether the seasonal

Figure 4.3: Year-on-year percent change in GDP

GDP: Value at 2000 prices
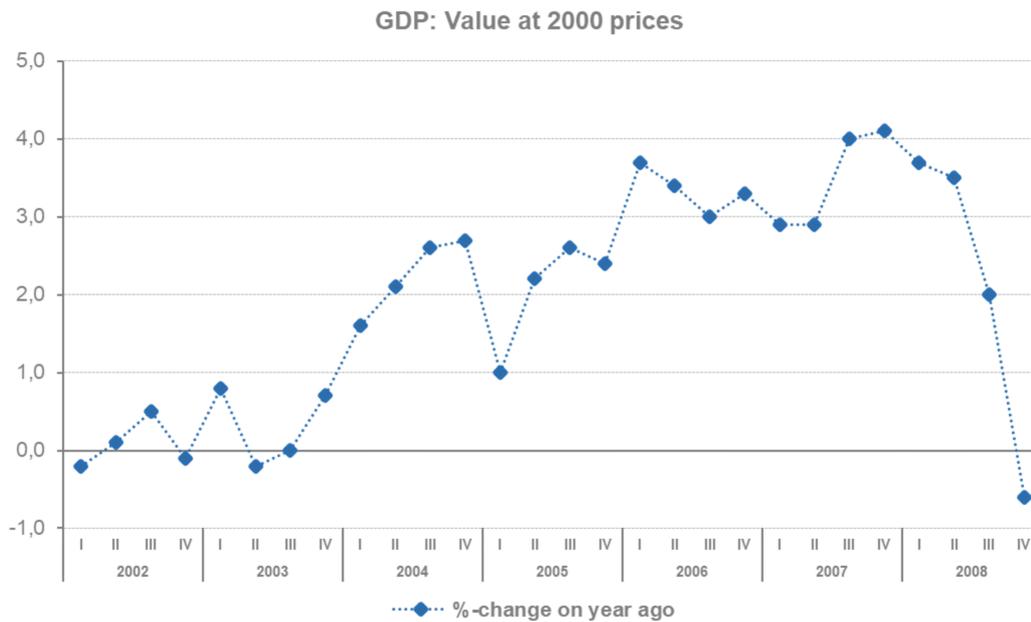
Source: Authors' own calculations

movements were relatively small compared to the other movements in the series. This problem was especially present in macro-economic series and for confidence indicators (consumer and producer confidence indices), and to a lesser extent in labour market series. See Figure 4.4 and Figure 4.5 for examples of this "masked" seasonal pattern in series of unemployment benefits and vacancies.

Figure 4.4: Unemployment benefits, example of a series where the seasonal pattern is "masked"

Unemployment benefits
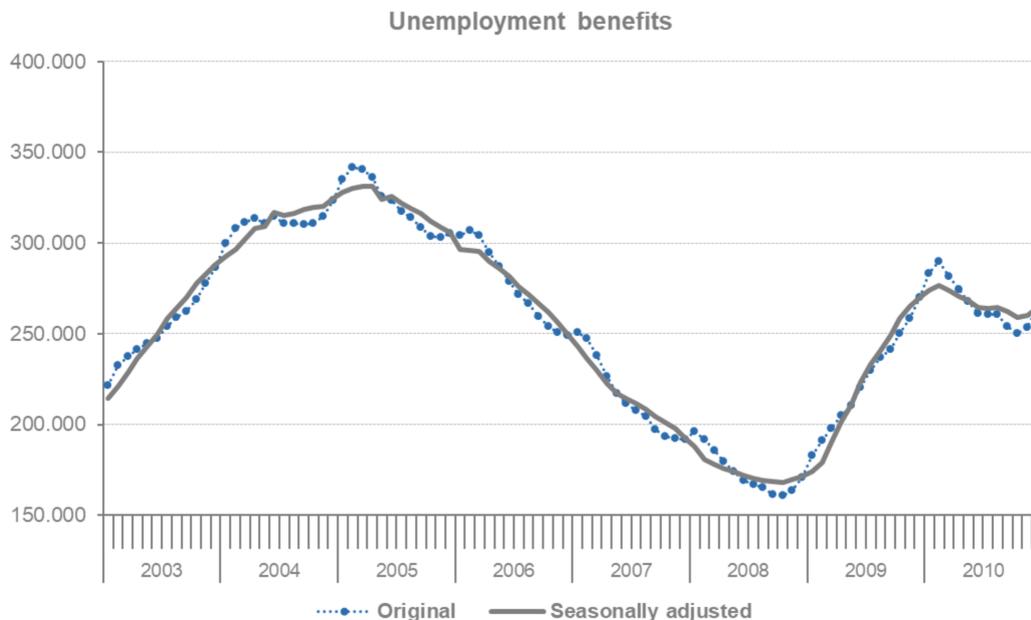
Source: Authors' own calculations

For both series, an almost linear increase in the original series can be seen in 2009, while the seasonally adjusted series hardly differs (as opposed to the previous years, where strong seasonal patterns can be seen). Later, the seasonal pattern returns, but still not as strong as before.

Figure 4.5: Vacancies, example of a series where the seasonal pattern is "masked"
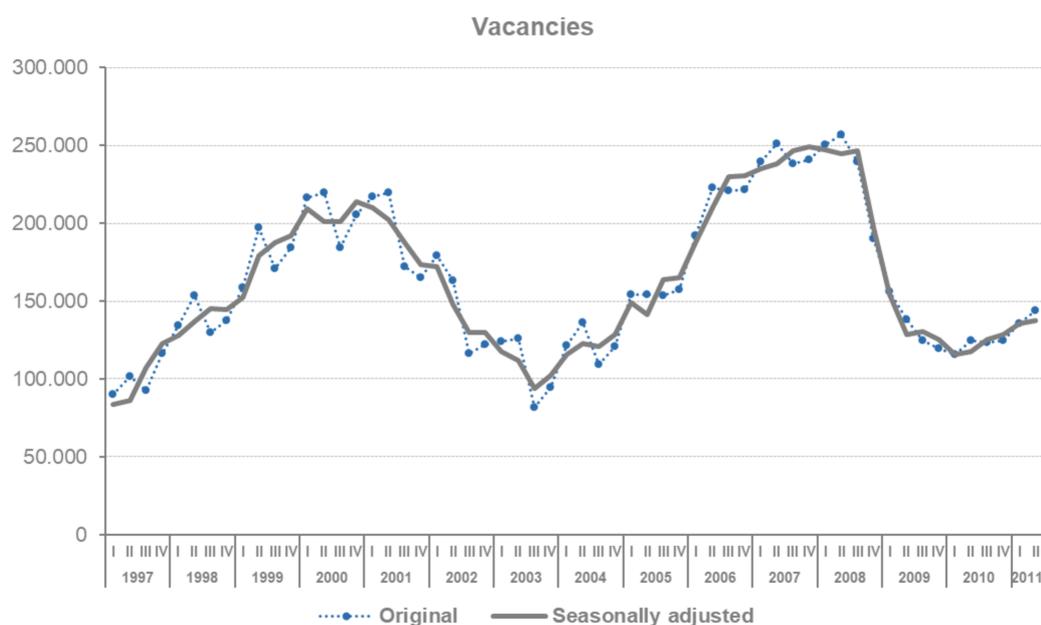
Source: Authors' own calculations

For quarterly series, a similar effect of the crisis on the seasonal pattern could not be seen. The seasonal pattern was clearly present at all times.

For both monthly and quarterly series a clear recession could be seen, with strong movements. These movements mostly did not seem to be structural, when series returned to levels (close to those) before the crisis. However, it remains to be seen if the effects of the crisis will completely disappear.

Influence on seasonal adjustment

For monthly series, the crisis actually made seasonal adjustment more difficult. This became clear from the substantial differences between the consecutive (monthly) estimates of the seasonal patterns. The main difficulties were:

1. Determining whether there actually was a crisis present in the series

2. Deciding whether to intervene in the seasonal adjustment process

3. The number of series that had to be published, and that could have been affected by the crisis

4. Possible consequences of an intervention on other steps of the production process

5. Other adaptations of the production process. For example, the production index was subject to a revision of the NACE classification at the same time of the crisis, so that the effects of these two events could not be separated. In this specific case, it was possible to recalculate the historical series based on this new classification, so that long time series for the new classification were available. Consequently, this made it possible to generate good estimates despite the crisis. If recalculating historical series would not be possible, a new classification could lead to difficulties.

For quarterly series, there were not many disturbances, as mentioned above, the seasonal was clearly visible (at least at a higher level of aggregation).

For the monthly series, most series are revised according to a concurrent adjustment policy, i.e. the data are kept fixed during the current year, and only updated once per year ("concurrent with annual review", see

Eurostat (2009; 2015)). For these series, the approach used before the crisis was continued during the crisis, i.e. a fine-tuned setup for each individual series, where outliers were detected automatically (i.e. if the data exceeds a critical value). Only for a small number of important series, a revision of the complete series was done every period.

Only part of the outliers was detected in an automatic fashion. Many monthly economic series fluctuated strongly and sometimes showed large peaks, leading to a case-by-case decision of what was an acceptable critical value for outlier detection. In these cases the standard critical value was lowered in steps of, for example, 0.1 (e.g., form 3.3 to 3.2), such that not too many nor too few outliers were detected for a reliable extraction of the seasonal pattern, or for acceptable adjustments at revisions, leading to plausible figures. This process of setting the critical value iteratively was done by seasonal adjustment experts, based on their experience.

The general strategy for the monthly series was to first wait and see how the series developed. After all, at first it is difficult to judge whether movements are temporary or more structural. A quick reaction could lead to a large adjustment later on. The development of the series and effects of seasonal adjustment procedures were monitored constantly, with a monthly decision on whether to intervene or not. These decisions were based on a combination of common sense (plausibility of results) and several quality indicators from X-12-ARIMA. Also, a possible change in parameters such as regression coefficients was closely monitored.

To prevent the need for strong adjustments, in these situations a relatively slow seasonal filter was preferred (slower than the standard X-12 filters). But even these sometimes proved not slow enough: the estimated seasonal patterns fluctuated more despite the slow filter. In Figure 4.6, monthly adjustments for the Industrial Production Index are shown. These adjustments were calculated by comparing the most recent series (June 2011 in this case) to first adjusted figures for each period. It can be seen that the adjustments fluctuate strongly in 2008 and 2009, and are very high several times. Table 4.1 shows the absolute average adjustment per year. Here as well, we can see that 2008 and 2009 contained substantially higher adjustments than surrounding years. In order to prevent these strong adjustments, it is recommended to build in more stability into the seasonal adjustment process in times of strong fluctuations.

Figure 4.6: Adjustments in case of concurrent seasonal adjustment
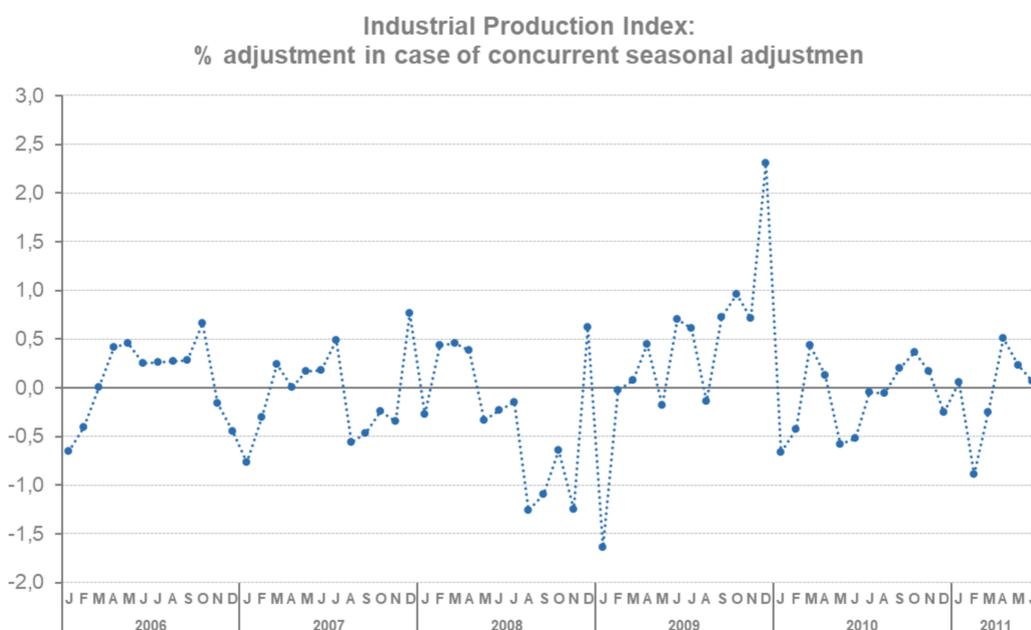
Source: Authors' own calculations

For several of these important series an intervention was considered necessary. In these cases, several outliers were set manually, after first waiting for several months. In Figure 4.7, an example of a series is

Table 4.1: Average absolute adjustment.

Year	Average absolute adjustment
2006	0.36
2007	0.38
2008	0.60
2009	0.71
2010	0.32
2011	0.33

shown where two manually set outliers have a strong influence on the Industrial Production Index. It was also considered whether the X-12-ARIMA parameters should be adjusted, which was done in several cases (for the regular annual review of setups).

For many quarterly series no specific adaptations to the seasonal adjustment process were done. Here, the usual approach was continued, including an automatic outlier detection (using default X-12-ARIMA settings). Since quarterly series are generally more stable than monthly series, automatic outlier detection did not give significant problems. For the highest aggregates, the crisis did not lead to more outliers than in other years, while for lower aggregates there were more outliers than usual. Although no significant adaptations were necessary, the series were more closely monitored during the crisis.

Figure 4.7: Example of a series with manually set outliers

Industrial Production Index
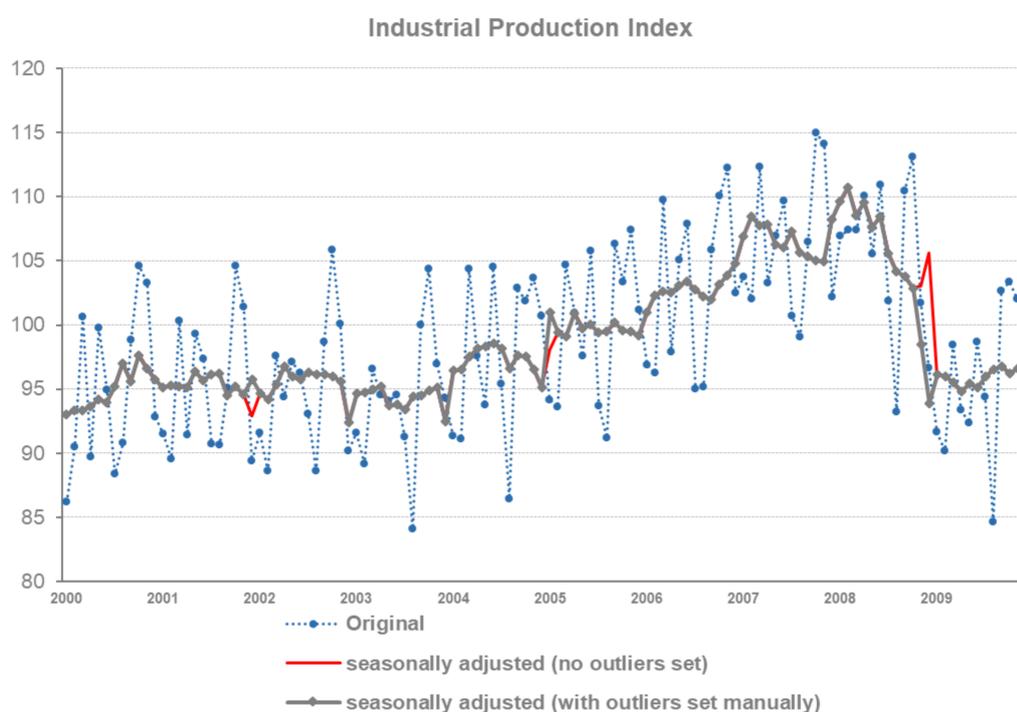
Source: Authors' own calculations

An important question that could not be answered yet is whether the seasonal patterns have in fact changed due to the crisis. They may have changed at least temporarily, but whether the changes are permanent is not clear. This may of course vary from series to series. For example, the seasonal pattern of the monthly series of turnover in hotels and restaurants has likely not changed significantly. Also for higher aggregates of quarterly series, no significant changes are expected, while for lower aggregates this is more likely. Knowing whether there was a change in pattern, is very useful. However, establishing this can only be done when the crisis is over.

Comparison between countries, conclusions and recommendations

Based on information from other European countries, a comparison could be made with their experiences, including the adaptations made to the seasonal adjustment process. The present approach (wait first, then set outliers carefully, using slow filters, or forecasting seasonal factors and then keeping these fixed for a while) is in accordance with international practice (although some countries adopted a more adaptive approach than others). Some countries chose to set more outliers than others, and in general, choices in setting outliers seem to be important. In order to quickly react to changing patterns, sometimes shorter series were used. In case of structural changes this may be a good approach.

The most important conclusion is that seasonal adjustment in times of an economic crisis requires specific and constant attention. Since developments in time series can differ substantially from usual developments, the standard seasonal adjustment procedures may not be applicable. It may be difficult or even impossible to quickly judge whether the developments seen over a short period of time are structural or only temporary changes. However, this aspect is crucial in deciding whether and how to adapt the seasonal adjustment process. Therefore, in these times of strong changes, we recommend a careful approach. Only if a sufficient number of observations was available to make a good judgement, a decision was taken to adjust the process or not. Most often a decision was made to manually set outliers.

Furthermore, it can be concluded that the crisis that started in 2008 has had strong effects on the seasonal adjustment of several series. For some important series (e.g. GDP) there is only a normal decline instead of a deep trough. This could be an explanation why the standard seasonal adjustment process led to acceptable results even in times of a crisis.

The seasonal adjustment process however, required more attention during this crisis than in other years, where it was not always clear whether and at what moment to intervene in the process. Besides, it was only possible to monitor the most important series, where common sense and the use of several indicators were used to decide to intervene or not. A single indicator that shows that there is an extraordinary situation, and that the seasonal adjustment process requires special attention, can have an added value. With such an indicator, more series can be monitored with less effort. Therefore, we recommend developing such an indicator. This indicator may use information from several series at the same time, since simultaneous extraordinary developments are an indication that it is likely that a structural change is taking place. Next to this indicator, recommendations should be developed on how to act when the indicator recommends an intervention.

General recommendations on all aspects of seasonal adjustment can be found in Eurostat (2009; 2015). There is a section on the treatment of outliers at the end of the series and at the beginning of a major economic change.

In publications, it can be desirable to keep outcomes of the seasonal adjustment process stable. More research is needed on how this can be done without ignoring real developments and reducing the size of adjustments. Eurostat (2009; 2015) gives recommendations on the revision strategy for seasonally adjusted data and the strategy adopted for the seasonal adjustment of long time series, since both can affect the stability and the path of the seasonally adjusted data.

At present it is not yet known whether and to what extent seasonal patterns have changed due to the crisis. For some series this may be the case, while for others it seems the pattern has not changed. For the cases where there was a change in pattern, the change is sometimes only temporary, and appeared during the most severe periods of the crisis. After that the old seasonal pattern then seems to return, as is the case for series where the seasonal pattern was "masked". We recommend to do more research on this when the crisis is over, in particular for all important series. Based on that, a decision can be made in what way the seasonal adjustment process should be adapted to deal with a future economic crisis.

Annex

4.A Modelling interrupted time series

Structural time series models

With a structural time series model, a series is decomposed in a trend component, seasonal component, other cyclic components, regression component and an irregular component. For each component a stochastic model is assumed. This allows the trend, seasonal, and cyclic component but also the regression coefficients to be time dependent. If necessary ARMA components can be added to capture the autocorrelation in the series beyond these structural components. See Harvey (1989) or Durbin and Koopman (2001) for details about structural time series modelling.

Consider a repeatedly conducted survey to produce series of official statistics for target variables. The variables of interest are defined as categorical variables measured on a nominal or an ordinal scale, means or totals. The population values of a categorical variable specify the distributions in the population over say K categories. For each variable a K-dimensional vector $y_t = y_{t,1}, \ldots, y_{t,k}$ is defined where the elements of y_t specify the proportions over the K categories. These K variables are subjected to the restriction

$$\sum_{k=1}^{K} y_{t,k} = 1, \text{ for all } t.$$

Population values for means and totals can be decomposed in say K subcategories and are defined as $y_t = (y_{t,+} y_{t,1}, \ldots y_{t,K})$. The first component, $y_{t,+}$, stands for the total and is broken down over K categories specified by the remaining estimates $\hat{y}_{t,k}$, $k = 1 \ldots K$. These $K+1$ variables are subjected to the restriction $y_{t,+} = \sum_{k=1}^{K} y_{t,k}$ for all t.

Based on the data observed in a sample, direct estimates for the unknown population values are obtained with design-based or model-assisted estimators, known from sampling theory. Examples are the Horvitz-Thompson estimator or the generalized regression estimator, Särndal et al. (1992). As a result, for each categorical variable K series are observed that specify the estimated proportions over K categories and are collected in the K-dimensional vector $\hat{y}_t = (\hat{y}_{t,1}, \ldots \hat{y}_{t,K})$, $t = 1, \ldots, T$. For variables defined as means or totals, $K+1$ series are observed that are collected in the $K+1$-dimensional vector $\hat{y}_t = (\hat{y}_{t,+}, \hat{y}_{t,1}, \ldots \hat{y}_{t,K})$, $t = 1, \ldots, T$. The sample estimates obey similar restrictions as their population values, that is $\Sigma_{k=1}^{K} \hat{y}_{t,k} = 1$ for proportions and $\hat{y}_{t,+} = \Sigma_{k=1}^{K} \hat{y}_{t,k}$ for means and totals.

Developing a time series model for survey estimates observed with a periodic survey starts with a model, which states that the survey estimate can be decomposed in the value of the population variable and a sampling error: $\hat{y}_{t,k} = y_{t,k} + e_{t,k}$, with $e_{t,k}$ the sampling error. Scott and Smith (1974) proposed to consider the true population value $y_{t,k}$ as the realisation of a stochastic process that can be properly described with a time series model.

In classical sampling theory, it is generally assumed that the observations obtained in the sample are true fixed values observed without error, see e.g. Cochran (1977). This assumption is not tenable if systematic differences are expected due to a redesign of the survey process. Van den Brakel and Renssen (2005) proposed a measurement error model for experiments embedded in sample surveys that links systematic differences between a finite population variable observed under different survey implementations. They consider

the observed population value obtained under a complete enumeration under two or more different implementations of the survey process as the sum of a true intrinsic value that is biased with a systematic effect induced by the survey design, i.e. $y_{t,k,l} = u_{t,k} + b_{k,l}$. Here $y_{t,k,l}$ is the population value of the k-th parameter at time t observed under the l-th survey approach, $u_{t,k}$ the true population value of this parameter and $b_{k,l}$ the measurement bias induced by the l-th survey process used to measure $u_{t,k}$. The systematic difference between two survey approaches is obtained by the contrast $y_{t,k,l} - y_{t,k,l'} = b_{k,l} - b_{k,l'} \equiv \beta_k$. In the case of embedded experiments, the systematic difference between two or more survey approaches is estimated as the contrast between estimates obtained from subsamples assigned to the different survey approaches. In the time series approach, these differences are estimated using an appropriate intervention variable. This allows for time dependent differences. For notational convenience, the subscript l will be omitted in $y_{t,k,l}$, since the survey approach will be indicated with the time period.

To keep the notation as parsimonious as possible, it is assumed that the autonomous development of the series of the indicator is modelled with a stochastic trend, a regression component and an irregular component. The regression component consists of an intervention variable with a time independent regression coefficient that describes the effect of the survey transition. This approach is initially proposed by Harvey and Durbin (1986). Seasonal, cyclic, ARMA, and other auxiliary regression components can be included in the model depending on the application at hand.

Based on the preceding considerations, the univariate structural time series model for the k-th component of $\hat{y}_t$ is defined as:

$$\hat{y}_{t,k} = L_{t,k} + \beta_k \delta_t + v_{t,k} + e_{t,k}, \tag{4.A.1}$$

with $L_{t,k}$ a stochastic trend, δ_k an intervention variable that describes under which survey the observations are obtained at period t, β_k the time independent regression coefficient for the intervention variable, $v_{t,k}$ an irregular component for the time series model of the population values $y_{t,k}$ and $e_{t,k}$ the sampling error. It is assumed that the irregular component is normally and independently distributed: $v_{t,k} \cong N(0, \sigma_v^2)$.

Surveys are often based on a rotating panel design. Such designs result in partially overlapping samples with correlated sampling errors. Particularly in these cases a separate component for the sampling error in the time series model might be required to capture this serial correlation. Through this component the estimated variances for the $\hat{y}_{t,k}$, which are generally available from the survey, can be included in the time series model as prior information. Binder and Dick (1990) proposed the following general form for the sampling error model to allow for non-homogeneous variance in the sampling errors:

$$e_{t,k} = \omega_{t,k} \tilde{e}_{t,k}, \tag{4.A.2}$$

where $\omega_{t,k}$ is the standard error of $\hat{y}_{t,k}$ and $\tilde{e}_{t,k}$ an ARMA process that models the serial correlation between the sampling errors. Abraham and Vijayan (1992), and Harvey and Chung (2000) applied MA models for the serial correlation in the sampling errors. Pfeffermann (1991), Pfeffermann et al. (1998) and Van den Brakel and Krieg (2009) used AR models for the serial correlation in the sampling errors. Autocorrelations can be estimated from the survey data and can be used, like the design variances of $\hat{y}_{t,k}$, as prior information in the sampling error model. Pfeffermann et al. (1998) developed a procedure to estimate the autocorrelation in the survey errors from the separate panel estimates of a rotating panel design and used this prior information to estimate the autocorrelation coefficients of an AR model.

If the observed series are based on non-overlapping cross-sectional samples, then there is no serial correlation between sampling errors. As a result the sampling error and the irregular component of the true population parameter cannot be separated. Therefore both terms are combined in one irregular term,

$$v_{t,k} + e_{t,k} = \varepsilon_{t,k}, \tag{4.A.3}$$

which is assumed to be normally and independently distributed with zero mean and a variance that is proportional to the variance of $\hat{y}_{t,k}$, i.e. $\varepsilon_{t,k} \cong N(0, \sigma^2_{\varepsilon,k} V\hat{a}r(\hat{y}_{t,k}))$, where $V\hat{a}r(\hat{y}_{t,k})$ is the estimated variance of $\hat{y}_{t,k}$. This variance structure still allows for non-homogeneous sampling variation, caused by differences in the yearly sample size, differences in the sampling design or differences in the variance of measurement error due to redesigns of the data collection procedures. The parameter $\sigma_{\varepsilon,k}$ is added, since the variation of $\varepsilon_{t,k}$ will not exactly follow the variance of the direct estimator. It is also assumed that the irregular components of equation 4.A.3 at different time points are uncorrelated: $Cov(\varepsilon_{t,k}\varepsilon_{t',k}) = 0$ for $t \neq t'$. As a result model in equation 4.A.1 simplifies to

$$y_{t,k} = L_{t,k} + \beta_k \delta_t + \varepsilon_{t,k}, \tag{4.A.4}$$

By combining $e_{t,k}$ and $v_{t,k}$ into one irregular term, it is implicitly assumed that the sampling error $e_{t,k}$ dominates the irregular term $\varepsilon_{t,k}$. The estimates for $\sigma^2_{\varepsilon,k}$ provide an indication to what extent the sampling error dominates the irregular term. They are expected to take values around one if the sampling error dominates the irregular of the population parameter, since the variance of $\varepsilon_{t,k}$ is taken proportional to the variance of $\hat{y}_{t,k}$. Estimates for $\sigma^2_{\varepsilon,k}$ larger than one indicate that $v_{t,k}$ contributes substantially to the variation of $\varepsilon_{t,k}$.

In many cases, there are no direct estimates available for the variance of $\hat{y}_{t,k}$. In such cases the variance can be approximated, for example by taking the variance inversely proportional to the sample size: $\varepsilon_{t,k} \cong N(0, \sigma^2_{\varepsilon,k}/n_t)$. This variance approximation implies that the only difference between the sample designs between the subsequent editions of the survey, is the sample size. In the case of categorical variables, the $\hat{y}_{t,k}$ are proportions, whose variances can be approximated with $V\hat{a}r(\hat{y}_{t,k}) = \hat{y}_{t,k}(1 - \hat{y}_{t,k})/n_t$.

For the stochastic trend, the so-called smooth trend model is assumed as an example since this trend model is widely applied in econometric time series analysis, see e.g. Durbin and Koopman (2001). In each particular application, other trend models should be considered. The smooth trend model is defined as:

$$
\begin{aligned}
L_{t,k} &= L_{t-1,k} + R_{t-1,k}, & (4.A.5)\\
R_{t,k} &= R_{t-1,k} + \eta_{t,R,k} & (4.A.6)
\end{aligned}
$$

with $L_{t,k}$ the level component and $R_{t,k}$ the stochastic slope component of the trend, and $\eta_{t,R,k}$ an irregular component. The smooth trend model in equation 4.A.6 is a special case of the local linear trend model, which also has an irregular term for $L_{t,k}$, see e.g. Durbin and Koopman (2001) equation 4.A.2, which can be considered as an alternative. It is assumed that the irregular components of equation 4.A.6 are normally and independently distributed, i.e. $\eta_{t,R,k} \cong N(0, \sigma^2_{R,k})$, and that they are uncorrelated at different time points, i.e. $Cov(\eta_{t,R,k}\eta_{t',R,k})$ for $t \neq t'$. Furthermore, it is assumed that the irregular components of equation 4.A.4 and equation 4.A.6 are uncorrelated: $Cov(\varepsilon_{t,k}\eta_{t',R,k})$ for all t and t'.

The intervention variable models the effect of the survey redesign. Three types of interventions are discussed: a level shift, a slope intervention, and an intervention of a seasonal pattern. Let T_R denote the time period at which the survey process is redesigned. In the case of a level intervention it is assumed that the magnitude of the discontinuity due to the survey redesign is constant over time. In this case δ_t is defined as a dummy variable:

$$\delta_t = \begin{cases} 0 & \text{if } t < T_R \\ 1 & \text{if } t \geq T_R \end{cases} \tag{4.A.7}$$

In the case of a slope intervention it is assumed that the magnitude of the discontinuity increases over time. This is accomplished by defining δ_t as:

$$\delta_t = \begin{cases} 0 & \text{if } t < T_R \\ 1 + t - T_R & \text{if } t \geq T_R \end{cases} \qquad (4.A.8)$$

It is also possible to define an intervention on the seasonal or cyclic pattern. Such interventions can be considered if an interaction is expected between the survey redesign and the months or the quarters of the year. In this case, a stochastic seasonal component is added to equation (1) or (4). Widely applied models are trigonometric models and the dummy variable seasonal model (see Durbin and Koopman (2001), section 3.2). Furthermore the intervention variable δ_t has the form (equation 4.A.7) and the regression coefficient β_k is replaced by a time independent seasonal component.

The interventions described so far assume that the redesign only affects the point estimates of the survey. A survey redesign could, however, also affect the variance of the measurement errors. An increase or decrease of the variance of the measurement errors will be reflected in the estimated variance of $\hat{y}_{t,k}$. A straightforward way to account for such effects is to incorporate the estimated variances of the survey estimates as prior information using sampling error model (equation 4.A.2). Another possibility is to define separate model variances for the irregular term $\hat{\varepsilon}_{t,k}$ in the measurement equation for the period before and after the implementation of the survey redesign, i.e. $Var(\varepsilon_{t,k}) = \sigma_{s,k.1}^2$ if $t < T_R$ and $Var(\varepsilon_{t,k}) = \sigma_{s,k.2}^2$ if $t > T_R$. The ratio between $\sigma_{s,k.1}^2$ and $\sigma_{s,k.2}^2$ can be used to test hypotheses about the equivalence of both variance components. This approach, however, requires a sufficient number of observations under both surveys to test the equivalence of these variance components with sufficient power.

The discontinuity in the series is modelled with an intervention variable that describes the moment that the survey process is redesigned. This approach assumes that the other components of the time series model approximate the real development of the population variable reasonably well and that there is no structural change in e.g. the trend or the seasonal component at the moment that the new survey is implemented. If a change in the real development of the population variable exactly coincides with the implementation of the new survey, then the model will wrongly assign this effect to the intervention variable which is intended to describe the redesign effect. Information available from series of correlated variables can be used to evaluate the assumption that there is no structural change in the real evolution of the parameter. Such auxiliary series can also be added as a regression component to the model, with the purpose to reduce the risk that a structural change in the evolution of the series of the target parameter is wrongly assigned to the intervention variable. An auxiliary series can also be included as a dependent variable in a multivariate model, which accounts for the correlation between the parameters of the trend and seasonal components, Harvey and Chung (2000) and Van den Brakel and Krieg (2012).

The risk that the intervention variable wrongfully absorbs a part of the development of the real population value can be reduced by applying parsimonious intervention parameters. Therefore, time dependent interventions, like an intervention on a seasonal component, must be applied carefully. These intervention parameters are more flexible and will easily absorb a part of the real evolution of the population value, particularly if only a limited number of observations after the survey change-over are available.

The intervention approach can be generalized in a straightforward way to situations were the survey process has been redesigned at two or more occasions. This is achieved by adding a separate intervention variable for each time that the survey process has been modified.

State-space representation

The structural time series models developed in subsection 2.3 for the separate parameters $\hat{y}_{t,k}$ of the vector $\hat{\boldsymbol{y}}_t$ comprise a H-dimensional structural time series model, where $H = K$ in the case of a K-dimensional categorical variable and $H = K + 1$ in the case of means and totals including a subdivision in K categories. The general way to proceed is to put this model in state-space representation and analyse the model with the Kalman filter. The state-space representation for this K-dimensional structural time series model reads as:

$$\hat{\boldsymbol{y}}_t = \boldsymbol{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t \qquad (4.A.9)$$

$$\boldsymbol{\alpha}_t = \boldsymbol{T} \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \qquad (4.A.10)$$

The measurement equation 4.A.9 describes how the observed series depends on a vector of unobserved state variables $\boldsymbol{\alpha}_t$ and a vector with disturbances $\boldsymbol{\varepsilon}_t$. The state vector contains the level and slope components of the trend models and the regression coefficients of the intervention variables. The transition equation 4.A.10 describes how these state variables evolve over time. The vector $\boldsymbol{\eta}_t$ contains the disturbances of the assumed first-order Markov processes of the state variables. The matrices in equations 4.A.9 and 4.A.10 are given by

$$\boldsymbol{\alpha}_t = (L_{t,1}, R_{t,1}, \ldots, L_{t,H}, R_{t,H}, \beta_1, \ldots, \beta_H)^T, \qquad (4.A.11)$$

$$\boldsymbol{Z}_t = (\boldsymbol{I}_{[H]} \otimes (1,0) | \delta_t \boldsymbol{I}_{[H]}), \qquad (4.A.12)$$

$$\boldsymbol{T} = Blockdiag(\boldsymbol{T}_{tr}, \boldsymbol{I}_{[H]}), \qquad (4.A.13)$$

$$\boldsymbol{T}_{tr} = \boldsymbol{I}_{[H]} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad (4.A.14)$$

with $\boldsymbol{I}_{[H]}$ the $H \times H$ identity matrix. The disturbance vectors are defined as

$$\boldsymbol{\varepsilon}_t = (\varepsilon_{t,1}, \ldots, \varepsilon_{t,H})^T, \qquad (4.A.15)$$

and

$$\boldsymbol{\eta}_t = (0, \eta_{t,R,1}, \ldots, 0, \eta_{t,R,H}, \boldsymbol{0}_{[H]}^T)^T. \qquad (4.A.16)$$

It is assumed that

$$E(\boldsymbol{\varepsilon}_t) = \boldsymbol{0}_{[H]}, Cov(\boldsymbol{\varepsilon}_t) = Diag(V\hat{a}r(\hat{y}_{t,1})\sigma_{\varepsilon,1}^2, \ldots, V\hat{a}r(\hat{y}_{t,H})\sigma_{\varepsilon,H}^2), \qquad (4.A.17)$$

and

$$E(\boldsymbol{\eta}_t) = \boldsymbol{0}_{[3H]}, Cov(\boldsymbol{\eta}_t) = Diag(0, \sigma_{R,1}^2, \ldots, 0, \sigma_{R,H}^2, \boldsymbol{0}_{[H]}^T), \qquad (4.A.18)$$

with $\boldsymbol{0}_{[p]}$ a column vector of order p with each element equal to zero. In the case that each measurement equation and each transition equation has its own separate hyperparameter, then equation 4.A.11 is a set of H univariate structural time series models. If the measurement equations or the transition equations share common hyperparameters, then equation 4.A.11 is a H dimensional seemingly unrelated multivariate structural time series model. This is for example the case if $\sigma_{\varepsilon,1}^2 = \ldots = \sigma_{\varepsilon,H}^2 = \sigma_\varepsilon^2$. The current state-space representation assumes a diagonal covariance structure for the measurement equation. A further extension is to allow for non-diagonal covariance structures, estimated from the survey data. For example, in the case of categorical data, the design covariances between the categories can be approximated with $Cov(\hat{y}_{t,k}\hat{y}_{t,k'}) = -\hat{y}_{t,k}\hat{y}_{t,k'}/n_t$. See Särndal et al. (1992), chapter 5, for more details about design covariances between survey variables.

The time independent regression coefficients of the intervention variables are also included in the state vector (as described by Durbin and Koopman (2001), subsection 6.2.2). The Kalman filter can be applied straightforwardly to obtain estimates for the regression coefficients. An alternative approach of estimating the regression coefficients is by augmentation of the Kalman filter, see Durbin and Koopman (2001) for details.

In the case of a categorical variable, each element of $\hat{\boldsymbol{y}}_t$ specifies the proportions over K categories. In other words, each variable makes up a K-dimensional series, which obeys the restriction that at each point in time these series add up to one, i.e. $\Sigma_{k=1}^{K} \hat{y}_{t,k} = 1$ and $0 \leq \hat{y}_{t,k} \leq 1$. As a result, the K regression coefficients of the intervention variables must obey the restriction $\Sigma_{k=1}^{K} \beta_k = 0$. The multivariate structural time series model in equation 4.A.11 can be augmented with this restriction by using the following design matrix in the transition

equation 4.A.10:

$$T = \text{Blockdiag}(T_{tr}, T_{iv}) \tag{4.A.19}$$

where T_{tr} is defined by equation 4.A.19 and

$$T_{iv} = \begin{pmatrix} I_{[K-1]} & 0_{[K-1]} \\ -1_{[K-1]}^T & 0 \end{pmatrix} \tag{4.A.20}$$

with $1_{[p]}$ a column vector of order p with each element equal to one. Due to T_{iv}, defined in equation 4.A.20, the regression coefficients as well as their Kalman-filter estimates obey the restriction $\Sigma_{k=1}^K \beta_k = 0$.

In the case of a population mean or total, the first element of $\hat{y}_t$ specifies the estimate for a variable and the remaining K elements a subdivision over K categories. These K+1 variables are subject to the restriction $\hat{y}_{t,+} = \Sigma_{k=1}^K \hat{y}_{t,k}$ for all t. As a result, the K+1 regression coefficients of the intervention variables must obey the restriction $\beta_+ = \Sigma_{k=1}^K \beta_k$. This restriction is obeyed if the design matrix in equation 4.A.21 is used in the transition 4.A.19 where

$$T_{iv} = \begin{pmatrix} 0 & 1_K^T \\ 0_K^T & I_K \end{pmatrix} \tag{4.A.21}$$

An intervention on a seasonal component can be implemented in a way similar to a level intervention. Let s denote the number of time periods of the seasonal set. The state vector α_t is augmented with $H \times s$ state variables to model the seasonal pattern for each parameter $\hat{y}_{t,k}$. The H regression coefficients β_k are replaced by another set of $H \times s$ state variables to model the intervention on the seasonal pattern for each target parameter. The design matrix of the measurement equation Z_t is augmented with a term $I_{[H]} \otimes z_{[s]}^T$, where $z_{[s]}$ is an s-dimensional vector that describes the relation between the observed series and the state variable of the trigonometric seasonal model or the dummy variable seasonal model. Furthermore $\delta_t I_{[H]}$ in Z_t is replaced by $\delta I_{[H]} \otimes z_{[s]}^T$. The design matrix of the transition equation is augmented with a block diagonal element $I_{[H]} \otimes T_s$, where T_s denotes the transitional relation for a trigonometric model or the dummy variable seasonal model. See Durbin and Koopman (2001), subsection 3.2 for expressions of $z_{[s]}$ and T_s. To force that the sum over the seasonal intervention variables of the K parameters equals zero, in the case of a categorical variable, the design matrix of the transition equation is augmented with $T_{iv} \otimes T_s$, where T_{iv} is defined by (3.10f). To impose the restriction for variables defined as totals and means, the design matrix of the transition equation is augmented with $T_{iv} \otimes T_s$, where T_{iv} is defined by equation 4.A.21.

Kalman filter

After having expressed the multivariate structural time series model in state-space representation and under the assumption of normally distributed error terms, the Kalman filter can be applied to obtain optimal estimates for the state variables as well as the measurement equation see e.g. Durbin and Koopman (2001). Estimates for state variables for period t based on the information available up to and including period t are referred to as the filtered estimates. The filtered estimates of past state vectors can be updated if new data become available. This procedure is referred to as smoothing and results in smoothed estimates that are based on the completely observed time series. So the smoothed estimate for the state vector for period t also accounts for the information made available after time period t. A widely applied procedure to smooth filtered point estimates and standard errors for the state variables is the fixed interval smoother. See Harvey (1989) or Durbin and Koopman (2001) for technical details. The non-stationary state variables are initialized with a diffuse prior, i.e. the expectations of the initial states are equal to zero and the initial covariance matrix of the states is diagonal with large diagonal elements. The time independent regression coefficients of the intervention variables are also initialized with a diffuse prior, as described by Durbin and Koopman (2001), subsection 6.2.2.

The use of a diffuse prior for the regression coefficients of the intervention variables is appropriate if no additional information about the size of discontinuities is available. As explained in subsection 2.2, external information about the size of the discontinuities can be available through the conduction of a parallel run

or through recalculation of existing data. This information can be used in the time series models by using an informative prior for the initialization of the regression coefficients. This can be done by using the direct estimate for the discontinuity obtained from the parallel run or through recalculation in the initial state vector for the regression coefficients and the estimated variance of this direct estimate as an uncertainty measure in the covariance matrix of the initial state vector. Another possibility to use the direct estimate of the discontinuities as prior information in the time series model, is to assume that the regression coefficient for the intervention are equal to the observed discontinuity in the parallel run. In this case the direct estimate for the discontinuity is treated as if it is a fixed value, known in advance. The uncertainty of using a survey estimate for the discontinuity is ignored.

Software

Several general statistical software packages contain procedures for the analysis of state-space models, e.g. SAS, R, and Eviews. There are, however, also two software packages that are specialized on the analysis of state-space models. A menu driven and user friendly package is STAMP where most common structural time series models are implemented (Koopman et al. (2009)). This package also allows the inclusion of interventions in structural time series models. The additional restrictions on the regression coefficients of the intervention variables in multivariate models, developed in subsection 2.4, are not supported. This kind of models can be implemented and estimated with Ssfpack 3.0 in combination with OxMetrics, see Doornik (2009) and Koopman et al. (2008). Ssfpack is a library of subroutines that can be used in a very flexible way to implement virtually all possible state-space models. It is less user friendly than STAMP and intended for the more experienced user of state-space models. Strong advantages of Ssfpack are that most of the advanced methods documented in Durbin and Koopman (2001) are implemented and that it handles the large sparse matrices, which are characteristically for multivariate state-space models, in an efficient way. This avoids many numerical problems, generally encountered by the implementation in programming environments of the aforementioned more general software packages.

For those who prefer a RegArima approach there is X13Arima-Seats (Bureau of the Census (2013)), Tramo-Seats (Caporello and Maravall (2004)) and Demetra+ (Grudkowska (2013)).

Bibliography

Abraham B., Vijayan K. (1992), Time series analysis for repeated surveys, Communications in Statistics: Simulation Computation, 21, 893-908.

Bell W.R., Hillmer S.C. (1990), The time series approach to estimation of periodic surveys, Survey Methodology, 16, pp. 195-215.

Binder D.A., Dick J.P. (1989), Modeling and estimation for repeated surveys, Survey Methodology, 15, pp. 29-45.

Binder D.A., Dick J.P. (1990), A method for the analysis of seasonal ARIMA models, Survey Methodology, 16, pp. 239-253.

Blight B.J.N., Scott A.J. (1973), A stochastic model for repeated surveys, Journal of the Royal Statistical Society, Series B, 35, 61-66.

Bondt G.J. de, Dieden H.C., Muzikarova S., Vincze I. (2013), Introducing the ECB Indicator on Euro Area Industrial New Orders, European Central Bank, Occasional Paper Series 149, http://www.ecb.europa.eu.

Box G.E.P., Tiao G.C. (1975), Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70, p. 70-79.

Brueton A. (1999), The development of Chain-Linked and Harmonised Estimates of GDP at Constant Prices, Economic Trends 552, pp. 39-45.

Bureau of the Census (2013), X-13 ARIMA-SEATS Reference Manual, Time Series Research Staff, Statistical Research Division, U.S. Census Bureau, Washington http://www.census.gov/srd/www/x13as/.

Caporello G., Maravall A. (2004), Program TSW, Tramo Seats Windows, revised reference manual, Banco de Espana, Madrid.

Chow G.C., Lin A.-L. (1971), Best linear unbiased interpolation, distribution and extrapolation of time series by related series. Review of Economics and Statistics 53, pp. 372-375.

Cochran W.G. (1977), Sampling techniques, Third edition. New York: John Wiley & Sons.

Conference Board (2001), Business Cycle Indicators Handbook, New York, NY: The Conference Board, http://www.conference-board.org/publications/publicationdetail.cfm?publicationid=852.

Daas P.J.H., Puts M.J.H. (2014), Big Data as a Source of Statistical Information, Paper written for the ECB workshop on using Big Data for forecasting and statistics, Frankfurt am Main, 7-8 April 2014.

Denton F.T. (1971), Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization, Journal of the American Statistical Association 66, pp. 99-102.

Doornik J.A. (2009), An Object-oriented Matrix Programming Language Ox 6, London: Timberlake Consultants Press

Durbin J., Koopman S.J. (2001), Time series analysis by state-space methods, Oxford: Oxford University Press.

Durbin J., Quenneville B. (1997), Benchmarking by state-space models. International Statistical Review, 65, pp. 23-48.

Eurostat (2001), Handbook on price and volume measures in National Accounts, http://ec.europa.eu/eurostat/en/web/product-manuals-and-guidelines/-/KS-41-01-543.

Eurostat (2009), Overview of international system of classifications, available online at:http://ec.europa.eu/eurostat/statistics-explained/index.php/File:International_system_of_classifications.PNGand: http://ec.europa.eu/eurostat/en/web/products-manuals-andguidelines/-/CA-26-99-069.

Eurostat (2009; 2015), ESS guidelines on seasonal adjustment, http://ec.europa.eu/eurostat/documents/3859598/5910549/KS-RA-09-006-EN.PDF.

Eurostat (2013), Weighting and Estimation -11-M-Preliminary Estimation Model-Based v3.2p, In: Memobust, www.cros-portal.eu/content/small-area-estimation.

Feder M. (2001), Time series analysis of repeated surveys: the state-space approach, Statistica Neerlandica, 55, pp. 182-199.

Findley D.F., Monsell B.C., Bell W.R., Otto M.C., Chen B.C. (1998), New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program, Journal of Business and Economic Statistics 16, pp. 127-152.

Foroni C., Marcellino M. (2013), A survey of econometric methods for mixed-frequency data, Norges Bank, Working paper 2013/06.

Grudkowska S. (2013), Demetra+ user manual. National Bank of Poland.

Harvey A.C. (1989), Forecasting, structural time series models and the Kalman filter, Cambridge: Cambridge University Press

Harvey A. C., Chung C.H. (2000), Estimating the underlying change in unemployment in the UK. Journal of the Royal Statistical Society, Series A, 163, pp. 303-339.

Harvey A.C., Durbin J. (1986), The effects of seat belt legislation on British road casualties: a case study in structural time series modelling, Journal of the Royal Statistical Society, Series A, 149, pp. 187-227.

International Labor Organization, (2003), Consumer Price Indices: An ILO Manual. International Labor Organization: Geneva

International Monetary Fund (2010), Producer price index manual, http://www.imf.org/external/pubs/ft/ppi/2010/manual/ppi.pdf.

Koopman S.J., Harvey A.C., Doornik J.A., Shephard N. (2009), Structural Time Series Analyser, Modeller and Predictor: STAMP 8.2: London: Timberlake Consultants Press.

Koopman S.J., Shephard N., Doornik J.A. (2008), SsfPack 3.0: Statistical algorithms for models in state-space form, London: Timberlake Consultants Press.

Krieg S., Blaess V., Smeets M. (2012), Small area estimation of turnover of the Structural Business Survey, Statistics Netherlands discussion paper 2012 03, http://www.cbs.nl/en-GB/menu/methoden/onderzoek-methoden/discussionpapers/archief/2012.

Lind J.T. (2005), Repeated surveys and the Kalman filter, Econometrics Journal, 8, 418-427.

McGuckin R.H., Ozyildirim A., Zarnowitz V. (2007), A more Timely and Useful Index of Leading Indicators, Journal of Business & Economic Statistics 25, pp. 110-120.

Organisation for Economic Co-operation and Development OECD, (2008), Handbook on constructing composite estimators: methodology and user guide, http://www.oecd.org/std/42495745.pdf.

Ostrom C.W. (1978), Time series analysis, Sage University paper series 9.

Pfeffermann D. (1991), Estimation and seasonal adjustment of population means using data from repeated surveys, Journal of Business & Economic Statistics, 9, pp. 163-175.

Pfeffermann D., Bleuer S.R. (1993), Robust joint modelling of labour force series of small areas. Survey Methodology, 19, pp. 149-163.

Pfeffermann D., Burck L. (1990), Robust small area estimation combining time series and cross-sectional data, Survey Methodology, 16, pp. 217-237.

Pfeffermann D., Feder M., Signorelli D. (1998), Estimation of Autocorrelations of Survey Errors with Application to Trend Estimation in Small Areas, Journal of Business & Economic Statistics, 16, pp. 339-348.

Pfeffermann D., Tiller R. (2006), Small area estimation with state-space models subject to benchmark constraints. Journal of the American Statistical Association, 101, pp. 1387-1397.

Rao J.N.K., Yu M. (1994), Small area estimation by combining time series and cross-sectional data, Canadian Journal of Statistics, 22, pp. 511-528.

Särndal C-E., Swensson B., Wretman J. (1992), Model Assisted Survey Sampling. New York: Springer Verlag.

Sartore M., Caporin D. (2006), Methodological Aspects of Time Series Back-Calculation, Working paper 56 Ca' Foscari University of Venice, Dipartimento Scienze Economiche.

Scott A.J., Smith T.M.F. (1974), Analysis of repeated surveys using time series methods, Journal of the American Statistical Association, 69, 674-678.

Scott A.J., Smith T.M.F., Jones R.G. (1977), The application of time series methods to the analysis of repeated surveys, International Statistical Review, 45, pp. 13-28.

Schouten B., Shlomo N., Skinner C. (2011), Indicators for Monitoring and Improving Representativeness of Response, Journal of Official Statistics 27, pp. 1-24.

Statistics Estonia (2009), Frequently asked questions about the chain-linking method, www.stat.ee/dokumendid/29861.

Tam S.M. (1987), Analysis of repeated surveys using a dynamic linear model, International Statistical Review, 55, pp. 63-73.

Tiller R.B. (1992), Time series modelling of sample survey data from the U.S. current population survey, Journal of Official Statistics, 8, pp. 149-166.

United Nations Statistical Division UNSD (2010), Manual on Statistics of International Trade in Services, MSITS 2002: http://unstats.un.org/unsd/pubs/gesgrid.asp?id=457 or http://ec.europa.eu/eurostat/documents/3859598/5855477/MANUAL-EN.PDF; concept MSITS 2010: http://unstats.un.org/unsd/statcom/doc10/BG-MSITS2010.pdf.

Van den Brakel J.A. (2008), Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey, Journal of the Royal Statistical Society, Series A, 171, 581-613.

Van den Brakel J.A. (2010), Sampling and estimation techniques for the implementation of new classification systems: the change-over from NACE Rev. 1.1 to NACE Rev. 2 in business surveys. Survey Research Methods. 4, pp. 103-119.

Van den Brakel J.A., van Berkel C.A.M. (2002), A design-based analysis procedure for two-treatment experiments embedded in sample surveys. An application in the Dutch Labour Force Survey. Journal of Official Statistics, vol. 18, pp. 217-231.

Van den Brakel J.A., Krieg S. (2009), Estimation of the monthly unemployment rate through structural time series modelling in a rotating panel design, Survey Methodology, vol. 35, pp. 177-190.

Van den Brakel J.A., Krieg S. (2012), Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design, Working paper, Statistics Netherlands, Heerlen.

Van den Brakel J.A., Renssen R.H. (2005), Analysis of experiments embedded in complex sampling designs, Survey Methodology, vol. 31, nr. 1, pp. 23-40.

Van den Brakel J.A., Roels J. (2010), Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. Annals of Applied Statistics, 4, pp. 1105-1138.

Van den Brakel J.A., Smith P.A., Compton S. (2008), Quality procedures for survey transitions - experiments, time series and discontinuities. Survey Research Methods. 2, pp. 123-141.

Van Delden, A. (2006), Diverging growth rates of supermarkets: causes and remedies (in Dutch) Statistics Netherlands research paper TMO-R&D-2006-02-09-ADLN.

Van Ruth F.J. (2002), The influence of the measurement period on consumer price indices, HCPI 02/398, paper presented at the Eurostat Working Party "Harmonisation of Consumer Price Indices", Luxemburg, 8-10 June 2002.

Van Ruth F., Schouten B., Wekker R. (2005), The Statistics Netherlands' Business Cycle Tracer, Methodological aspects; concept, cycle computation and indicator selection, http://www.cbs.nl/NR/rdonlyres/253FD272-B93E-46FF-A474-1E5A396C81F1/0/2005methodebusinesscycletracerart.pdf.

Van Velzen M., Wekker R., Ouwehand P. (2011), Seasonal adjustment, Statistics Netherlands Methods Series, The Hague, http://www.cbs.nl/NR/rdonlyres/75B108E9-FE3E-4506-904D-166AD9F785A6/0/2011x37062art.pdf.

Zhang L.C. (2012), Topics of statistical theory for register-based statistics, Statistica Neerlandica 66, pp. 41-63.

5

Review of Parametric and Non-Parametric Variable and Model Selection Techniques

Author

Dominique Guégan
Université Paris 1 Panthéon-Sorbonne
dominique.guegan@univ-paris1.fr

Table of Contents

5.1 Introduction

The chapter describes alternative parametric, semi-parametric and non-parametric approaches for variable and model selection. A comparative analysis of alternative techniques evidencing strong and weak points as well as advantages and draw backs is presented. Examples using Euro zone data sets and US data sets illustrate the interest of these various methodologies in order to build composite indicators for the study of the GDP.

5.2 Parametric modellings for variable and model selection

There is compelling empirical evidence that economic systems exhibit several regimes often created by jumps, conducing to switch from one regime to another. When such a switch occurs the distribution of the data changes. For example, the macro-economy periodically switches from expansion to recession and back again, and dynamics differ between these two regimes. To take these phenomena into account is fundamental for the modelling of business cycles and to build composite indicators to explain it. One way to model such transitions is to use extended Markov processes. If the AR regression model is a natural approach for economists, it is now well known that it cannot take into account breaks, switches or volatility and mismatches the modelling of the GDP, and can delay or not the detection of the entrance in recession periods. Nevertheless, the use of VAR modelling including exogenous variables permit to bypass the inconvenience of the univariate modelling.

The use of non-linear models can be also appropriated to catch some specific features of the data at certain dates: Markov switching models, Hamilton (1988) have shown their interest to recover recessions periods in the US economic zone; Threshold modellings including STAR and SETAR models can also have good performance, Ferrara and Guégan (2004), Teräsvirta (2006), Billio et al. (2009); Factors models is also a root to integrate kinds of non-linearities, Kim and Nelson (1999). When the choice to introduce other economic variables or financial variables is retained related-GARCH modellings can catch a part of the volatility of the data sets, Filardo (1994), Cai (1994), Hamilton and Susmel (1994), and Krolzig and Toro (2000), among others.

Keeping the idea that the economic activity in market economies is characterized by phases of upturns followed by phases of depression, it makes sense to represent the state of an economy by an index (the reference cycle) describing the common behavior of interest variables. Thus, any model will contain such an index. A general multivariate representation including all the previous modellings as special cases can be proposed. Thus if Y_t represents the interesting variable (GDP for instance) and X_t the explanatory variables which can be used as indicators, then we will have the following representation:

$$Y_t = C_{s_t} X_t + D_{s_t} \varepsilon_t, \ t = 1, \ldots, T, \qquad (5.1)$$

and

$$X_t = A_{s_t} X_{t-1} + B_{s_t} v_{t-1}, \qquad (5.2)$$

$(s_t)_t$ is a random variable, $X_t = (Y_t, Z_t)$, X_t will play the role of factor variable including the exogenous variables Z_t. The components which appear in (1) and (2) are orthogonal.

This model contains the VAR modelling, the Markov switching modelling the STAR and SETAR process, the Factor models and their state space representation, and also all the related-GARCH modellings.

- For linear regression with or without exogenous variables in univariate or multivariate approach we restrict to equation (1) where C_{s_t} does not depend on time t, and D_{s_t} =1.

- For static factor modellings, $A_{s_t}, B_{s_t}, C_{s_t}$, and D_{s_t}, will not depend on time t.

- For Markov switching modelling, we consider only the equation 5.1 in which s_t is a Markov chain and $D_{s_t} = 1$.

- For non-linear modellings like SETAR, STAR or related GARCH, we use again the equation (1) and the non linearity is inside C_{s_t} :

    1. For related SETAR modellings: $C_{s_t} = \Phi(B)1_{X_t < c} + \Theta(B)1_{X_t \geq c}$, where B is the back-shift operator and Φ and Θ are polynomials in B.

    2. For related GARCH modellings: C_{s_t} is constant, and D_{s_t} is a polynomial of any order in X_t

$$D_{s_t}\varepsilon_t = F_{s_t} + G_{s_t}X_t^\delta + H_{s_t}\varepsilon_t^\delta \tag{5.3}$$

with $\delta \in R$.

All these modellings have their interests and limits. The AR and VAR modellings are very common, nevertheless they do not take into account a lot of events and they can be used uniquely for very short term predictions. The indicators used in these modellings impact the reference business cycle only in a linear way. When there are a lot of variables, the fitting with VAR models can be limited because of estimation procedure whose robustness diminishes as soon as the number of variables increases. In that latter case the factor models will be privileged. In factor models when we have only one common factor affecting all the time series only contemporaneously such a factor can be interpreted as the reference cycle. They permit to add orthogonal measurement error and the economic model has a factor analytic structure, even if it appears simple in a first insight. Changing the assumptions on this common factor,we can make this model more complex.

For VAR modellings and factor modellings the assumptions of Gaussianity for the residuals is always considered, making the use of these classes of models very restrictive. Note also that in the Markov switching representation the processes $(s_t)_t$ and $(\varepsilon_t)_t$ are independent; $(\varepsilon_t)_t$ is a strong Gaussian white noise, and concerning the use of $(s_t)_t$ we need to know $P[s_0|I_0]$. Nevertheless the use of Markov chains permits to introduce some kinds of non-linearity. We are free of these kind of restrictions with the large class of STAR and SETAR modellings now well documented and whose economic indicators can be clearly identified when we switch from one state to another, permitting to build, in an easy way, the corresponding composite indicators.

As soon as the coefficients or matrices appearing inside the previous modellings are dependent of time t, some non-stationarity is included in the modelling itself. This important fact is generally not taken enough into account, Guégan (2009) for deeper discussion. Another possible extension is to introduce information on the expected duration of a state or a regime in the transition matrix. The relevant question will be: Given that we are currently in regime j or in state j ($s_t = j$), how long – on average – will regime j last? Answering this question would permit to know for how long the economy will be in a recession or a boom. If we consider the Filardo (1994) model, to answer to this question, we can consider a special form of the time-varying transition probabilities in a regime-switching model of the business cycle, replacing the exogenous variable by a duration of the current regime (boom or recession), Durland and McCurdy (1994) and Lam, Pok-Sang (2004) for possible discussion.

5.3 Non-Parametric Modellings

Most of the works inside the economic theory are based on econometric models using specified parametric modellings as previously recalled and likelihood based methods of inference. Under regularity conditions, maximum likelihood estimators consistently estimate the unknown parameters of the likelihood function.

Asymptotic theory on tests and estimation permits to provide consistent estimates for the parameters. However, when the parametric model is not true, these estimators may not be fully efficient, and in many cases – for example in regression when the functional form is mis-specified – may not even be consistent. The costs of imposing the strong restrictions required for parametric estimation and testing can be important. Therefore, much effort has gone into developing procedures that can be used in the absence of strong *a priori* assumptions, which are interesting alternatives to study the economic cycle.

In this section, we consider alternative approaches based on non-parametric smoothing methods which do not impose parametric restrictions on functional form in order to investigate the business cycle. There are several reasons for which we are interested by non-parametric techniques.

- First, these methods can be employed as a convenient and succinct mean of displaying the features of a data set.

- Second they can be used for diagnosis checking of an estimated parametric model.

- Third, we can use these methods to conduct inference under very weak assumptions

- Fourth, non-parametric estimators are frequently required in the construction of estimators used in semi-parametric modellings.

- Finally, the techniques like the k-Nearest Neighbors $(k - NN)$ and Radial Basis Functions (RBF) methods permit to model existence of non-linearity and non-stationarity inside the data sets.

In order to use these methods to re-build a signal as the economic cycle for instance, we consider the simple problem which consists to estimate the relationship between two random variables, say X and Y . Regression analysis is concerned with the question of how Y (the dependent variable) can be explained by X (the explanatory or regressor variable). This means the existence of a relationship of the form:

$$Y = m(X), \tag{5.1}$$

where $m(.)$ is a function in the mathematical sense. The random variables X, in the expression (4), can be any variables : exogenous random variables or lags of the variable of interest $Y_{t-1}, \ldots, Y_{t-p}$. The underlying principle that theoretical laws usually do not hold in every individual case but merely on average is considered here and we can formalize the expression (4) as

$$m(x) = E[Y|X = x]. \tag{5.2}$$

Given a finite set of observations (y_i, x_i), $i = 1, \ldots, n$, if we suppose that:

$$y_i = m(x_i) + \varepsilon_i, \ i = 1, \ldots, n, \tag{5.3}$$

where $(\varepsilon i)_i$ is a sequence of independent random errors satisfying $E[\varepsilon i|X_i = x] = 0$ and $Var[\varepsilon_i|X_i = x] = \sigma^2(x)$, equation (6) says that the relationship (4) does not need to hold exactly for the ith observation but is disturbed by the random variable ε. Yet, equation (5) says that the equation (4) holds in average, i.e. the expectation of Y on the condition that $X = x$ is given by $m(x)$. Thus given this information set, we can estimate $m(.)$ in the following way:

$$\hat{m}(x) = \sum_{i=1}^{n} w_{i,n}(x)Y_i, \tag{5.4}$$

where $w_{i,n}$ are weights to specify and estimate, and arise from different motivations and possess various statistical properties, Silverman (1986), Guégan (2003), Härdle et al. (2004). Were call three modellings for

$\hat{m}(.)$.

The so-called Nadaraya-Watson kernel estimate has the following expression:

$$\hat{m}_h = \frac{\sum_{i=1}^{n} K_h(x - x_i)y_i}{\sum_{i=1}^{n} K_h(x - x_i)} \tag{5.5}$$

where the kernel $K(.)$ is a piecewise continuous function, symmetric around zero, integrating to one which need not have bounded support, and is generally a positive probability density function. The bandwidth h determines the degree of smoothness of $\hat{m}_h(x)$. This can be immediately seen by considering the limits for h tending to zero or to infinity, respectively. Indeed, at an observation x_i, $\hat{m}_h(x_i) \longrightarrow y_i$, as $h \longrightarrow 0$, while at an arbitrary point x, $\hat{m}_h(x) \longrightarrow \bar{Y}$, as $h \longrightarrow \infty$. These two limit considerations make it clear that the smoothing parameter h in relation to the sample size n should not converge to zero too rapidly nor too slowly. Under such conditions, the kernel estimator is asymptotically Gaussian. The Nadaraya-Watson estimator has an obvious generalization to d-dimensional random variables and p^{th} order kernels. In that case, we need to assume a common bandwidth h fixed, then the asymptotic bias is $O(h^p)$, and p is an even integer, while the asymptotic variance is $O(n^{-1}h^{-d})$. The use of this estimate is limited due to these strong constraints on the choice of the bandwidth.

The k-nearest neighbors regression estimate is:

$$\hat{m}(x) = \sum_{i \in N(x)} w(X_t - X_{(i)})X_{(i)+1}, \tag{5.6}$$

where $w(.)$ is a weighting function associated to neighbors. We can distinguish three types of $k - NN$ regression estimates for which it is noteworthy that the parameter k needs to be estimated. The neighborhood $N(x)$ is defined through the variables X which are among the k-nearest neighbors of a point x, thus $N(x) = i : X_{(i)}, i = 1, \ldots, k$, where $X_{(i)}$ represents the k-nearest neighbors of x. We retain the following choices for the weighted function $w(.)$:

1. The first one corresponds to the so called uniformly weighted $k - NN$ regression:

$$\hat{m}_n(x) = \frac{1}{k} \sum_{i \in N(x)} X_{(i)+1} \ i.e. \ w(x - X_i) = \frac{1}{k}. \tag{5.7}$$

2. The second one is a generalization of the previous one by considering a non-uniform weight non-dependent on $(X_n)_n$:

$$\hat{m}_n(x) = \sum_{i \in N(x)} w_i X_{(i)+1}, w_i \in R, \ then \ w(x - X_i) = w_i. \tag{5.8}$$

3. The third one takes into account the distance of the point observation and the neighbors, and depends on $(X_n)_n$:

$$\hat{m}_n(x) = \frac{\sum_{i \in N(x)} K(\frac{x-X_i}{R(k)})X_{i+1}}{\sum_{i \in N(x)} K(\frac{x-X_i}{R(k)})} \tag{5.9}$$

where $R(k)$ corresponds to the distance between x and the further k-nearest neighbors and $K(.)$ is a weighting function. Estimating $m(.)$ using the $k - NN$ method at a point x, when the data are sparse might occur when the k-nearest neighbors are rather far away from x (and each other), consequently we end up with a wide neighborhood around x, for which an average of the corresponding values of $X(i)$ is computed.

The radial basis function regression estimate has the form:

$$\hat{m}_n(x) = \sum_{i=1}^{k} w_i \phi(||\underline{X}_n - c_i||, r_i), \tag{5.10}$$

where the radial basis function ϕ is an application defined from R^d to R^d and is characterized by its centroid c and its width r. This means the observations x_i are embedded in a space of dimension d in which the parameters which characterized this function will be estimated. The radial basis function can be chosen, for instance, among the following functions:

- Spline function : $\phi(x, r) = \frac{x^2}{r^2} log\left(\frac{x}{r}\right)$,

- Gaussian function : $\phi(x, r) = exp\left(-\frac{x^2}{2r^2}\right)$,

- Multiquadric function : $\phi(x, r) = \sqrt{x^2 + r^2}$,

- Inverse multiquadric : $\phi(x, r) = \frac{1}{\sqrt{x^2 r^2}}$.

In relationship 5.10, $\underline{X}_n = (X_n, X_{n-1}, \ldots, X_{n-(d-1)}) \in R^d$ and the parameters to estimate are c_i, r_i and w_i which is done using the k-means cluster method. k-means method corresponds to a partition of the individuals into clusters such that each individual belongs to the cluster whose center is closest in Euclidean distance. To determine all these parameters, we first embed the data set in a space of dimension d. As the embedding dimension d is not known a priori, we proceed step by step, beginning at $d = 1$, etc. Then given a d-dimensional space, we use a k-means clustering method to partition the points such that each one belongs to the cluster whose centre is the closest in the Euclidean distance sense (the procedure is done for a given function ϕ). At this step, the observations in R^d are organized around k clusters with their centres $c_i, i = 1, \ldots k$. Then, we estimate the width r_i using the r centers $c_j (r \leq k)$ which are closest to c_i, such that, $i = 1, \ldots, k$:

$$r_i = \frac{1}{r} \sqrt{\sum_{j=1}^{r} ||c_i - c_j||^2} \tag{5.11}$$

.

This choice permits to avoid that the clusters overlap. As soon as the function ϕ and the parameters $(c_i, r_i), i = 1, \ldots, k$ are known, then $\phi(||\underline{X}_n - c_i||, r_i)$ is known and the function $\hat{m}(x)$ is linear in w_i, so we can estimate w_i by ordinary least squares method. Finally the one-step-ahead value obtained with the RBF method is given by the relationship (13), where $\phi(x, r)$ is one of the previous radial basis function. Some references on radial basis functions are Wegman and Wright (1983) and Girosi and Anzelotti (1993).

Some comments: In most of the previous methods we do not discuss the stationary assumption on the given set of data. This assumption is not assumed with k-NN nearest neighbors method nor with radial basis functions method. It is usually assumed with the kernel method. In practice many time series have trends and are therefore non-stationary. These trends may be removed prior to an analysis of the stationary part of the process if the trend function is known. In most cases it is unknown. If we assume that the trend is a smooth deterministic function, we can use non-parametric techniques to remove it. If we assume that the trend is stochastic, we can remove it by differentiating d times the data set.

Semi parametric models offer a compromise between parametric modelling and the non parametric approaches. For instance, if it is necessary to account for both functional form and correlation of general nature, semi parametric models may be preferred. By semi parametric modellings we mean several approaches:

1. The data sets are modelled through a parametric model and the density of the observable data, conditional to an information set, is completely specified by a finite dimensional parameter θ and an unknown

function $F(.)$. Non-parametric methods can be developed to estimate the distribution function F. The exhaustive monograph of Bickel et al. (1992) presents a comprehensive theory on this approach.

2. Some parts of the parametric modellings can be estimated using non-parametric regression methods as proposed previously. This approach allows for general functional form, Engle et al. (1996).

3. In a parametric model, we can embed variables which have been estimated or forecast using non-parametric techniques, Ferrara et al. (2010).

5.4 Model and Variable selection

An interesting approach for variable and model selection, for instance in case of GDP modelling, is encompassing. Encompassing corresponds to a familiar notion in most sciences: an essential criterion in validating a new model or theory lies on its capability to account for (encompass) findings or failures of earlier models. While we will never know whether a statistical model is correct, we should do our best to test whether it could be refuted, as well as corroborate it with empirical and theoretical knowledge.

For instance if we consider a three variables model defined as

$$Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i, \tag{5.1}$$

where the triplets $(\varepsilon i, X_{2,i}, X_{3,i})$ are independent and $(\varepsilon i | X_{2,i}, X_{3,i}) = N(0, \sigma^2)$, there are three types of hypotheses of interest. The first hypothesis of interest is to restrict one of the parameters to be zero, for instance $\beta_3 = 0$. In that latter case we compute the log likelihood statistic $LR_{\beta_3} = 0$. A second hypothesis is to restrict the parameters of both the regressors to be zero, so $\beta_2 = \beta_3 = 0$. In that latter case it is the log likelihood statistic $LR_{\beta_2 = \beta_3 = 0}$ that we compute. We can remark that if the hypothesis $\beta_3 = 0$ is true, the second hypothesis $\beta_2 = \beta_3 = 0$ could be tested within a two variables model using the likelihood $LR_{\beta_2 = 0 | \beta_3 = 0}$, and then we have the relationship: $LR_{\beta_2 = \beta_3 = 0} = LR_{\beta_3 = 0} + LR_{\beta_2 = 0 | \beta_3 = 0}$.

This is conduced first using a $LR_{\beta_3 = 0}$ test, and if that is accepted, then testing $\beta_2 = 0$ using $LR_{\beta_2 = 0 | \beta_3 = 0}$. With this conditional interpretation a χ^2 distribution would apply for both tests. Thus, a multiple testing approach seems interesting but how will conducting a number of tests affect inference? In fact, the hypothesis $\beta_2 = 0$ would be considered only if testing supports setting β_3 to zero. All will depend on the value of $LR_{\beta_3 = 0}$. We can go seriously wrong by applying the χ^2 distribution for the second test when the $LR_{\beta_3 = 0}$ test is close to the critical value. Thus, various alternatives can be considered to test the good modelling (models' selection) or the good information set (variables' selection).

5.4.1 Encompassing test: a simple approach for linear modelling

The empirical econometric problem is that we do not have complete and correct stochastic theories from which to derive a "Correct" statistical model, so often we have to decide between different modellings, and encompassing models is a way to solve this problem.

We consider three models M_0, M_1, and M_2. The model M_0 is to be thought of as the general unrestricted model, and both M_1 and M_2 are special cases of M_0. For that consider three variables such as GDP, Y_i, interest rates, $X_{1,i}$ and oil $X_{2,i}$. For the purpose of the presentation, we assume that the variables $(Y_i, X_{1,i}, X_{2,i})$ satisfy a trivariate normal distribution. We assume also that the variables have zero expectations, so $E[Y_i] = E[X_{1,i}] = E[X_{2,i}] = 0$. The general model is denoted M_0, and is given by the conditional relation $E[Y_i | X_{1,i}, X_{1,i}]$ assumed linear, that is:

$$M_0 : Y_i = \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i \tag{5.2}$$

where $(\varepsilon i | X_{1,i}, X_{2,i})$ follows a Gaussian law $N(0, \sigma_\varepsilon^2)$, and $(\beta_1, \beta_2, \sigma_\varepsilon^2) \in R^2 \otimes R_+$ are constant parameters.

We first consider a competitive model representing $E[Y_i | X_{1,i}]$ that is:

$$M_1 : Y_i = \alpha X_{1,i} + \nu_i \tag{5.3}$$

.

with $(\nu_i | X_{1,i})$ follows a Gaussian law $N(0, \sigma_\nu^2)$. the model M_1 is nested in M_0, in that it can be found by imposing $\beta_2 = 0$ in (15).

In the general case where β_2 is unrestricted, the parameter α of M_1 will be a function of β_1 and β_2 of M_0. In that latter case the parameter α of the model M_1 can be derived from the parameters of the general model M_0, and then we derive how the latter model explains the former one. In such setting, knowledge of M_0 explains M_1, and we say that M_0 encompasses M_1.

We can consider now a second conditional model of Y_i denoted M_2, which postulates $E[Y_i | X_{2,i}]$ is:

$$M_2 : Y_i = \delta X_{2,i} + \eta_i \tag{5.4}$$

.

with $(\eta_i | X_{2,i})$ follows a Gaussian law $N(0, \sigma_\eta^2)$. Then we have $\delta = \beta_2 + \beta_1 b_{12}$ with $b_{12} = m_{12} m_{22}^{-1}$ where $m_{jk} = E[X_{j,i} X_{k,i}]$. The models M_1 and M_2 are said to be non-nested, in that neither model is a special case of the other, unless there is a perfect correlation between $X_{1,i}$ and $X_{2,i}$. Nevertheless we can be interested to verify if model M_1 can explain model M_2, which would say that M_1 encompasses M_2. Two possibilities arise: if $\beta_2 = 0$ then M_1 encompasses M_2, but in that latter case M_1 coincide with M_0, so both models can be used. Conversely when $\beta_2 = 0$, then M_1 should fail to explain the value δ found by M_2. Thus we can test the hypothesis that M_1 encompasses M_2 in two ways: either by testing: $\beta_2 = 0$ in the unrestricted model M_0, or by testing directly $\delta = \alpha b_{21}$ using the restricted models M_1 and M_2 only. These two assumptions are equivalent considering two types of tests. In the first set of tests, the models M_1 and M_2 are compared directly. The second test is based on testing the models M_1 and M_2 respectively, as restrictions of the joint model M_0. Remark that a necessary condition for M_1 encompasses M_2 is that M_1 fits better than M_2 when $\beta_1 \neq 0$, that is $\sigma_\eta^2 < \sigma_\nu^2$.

In summary, encompassing is a general principle wherein a given model is evaluating by deriving its implication for other existing models and checking that such implications are not rejected. This is a determinant, for instance, to decide what are the good economic indicators which explain the behavior of the GDP. Given two models M_1 and M_2, we compare these two models and their specifications, say can M_1 explain the reported results of M_2, or vice-versa? A way to deduce mis-specifications in M_2 more far-reaching approach is to use M_1; for example, if M_1 involved a variable for example $Z_{k,t}$ that experienced a large location shift in-sample, but M_2 excluded that variable, then M_1 should predict a structural break in M_2 at that time $Z_{k,t}$ shifted. If no such shift occurred in M_2, that is evidence against M_1, whereas if the predicted shift did occur, that is evidence in favor of M_1 and against M_2.

5.4.2 Criteria for model selection

When we have two parametrized competing models estimated by maximum likelihood, there exist popular procedures (called criteria) for model selection. They are derived from the Akaike Information Criteria. Consider a large number, say N candidate explanatory variables, where the general linear model is postulated to be:

$$Y_i = \sum_{i=1}^{N} \gamma_i Z_{i,t} + \varepsilon_i \tag{5.5}$$

when the conditional data-generating equation is in fact nested in 5.2 as:

$$Y_i = \sum_{i=1}^{n} \beta_j Z_{(j),t} + \nu_i, \tag{5.6}$$

where $\nu_t \sim N(0, \sigma_\nu^2)$ is independent of all the $Z_{i,t}$ in equation 5.2 for $\leq N$. The variables that actually matter in equation 5.3 are denoted by $Z_{(j),t}$ to reflect that, say, only the first and the third variables for instance of the original set might be relevant. The objective is to select these relevant regressors, namely the n different $Z_{(j),t}$ where $\beta_j = 0$ in (17), and to eliminate the irrelevant regressors, which are all the other $N - n$ variables, then all N are initially thought to be potentially relevant.

The information criteria called first FPE, Akaike (1969), Akaike (1973) and Akaike (1974) Final Prediction Error, penalizes the likelihood of the model of interest by $2n/T$ for n parameters (regressors here) and a sample size T. The AIC criteria, for model in equation 5.3 is:

$$AIC_n = ln\left(\hat{\sigma}_n^2\right) + \frac{2}{T}, \tag{5.7}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \sum_{i=1}^{n} \beta_j Z_{(j),t}^2) = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t. \tag{5.8}$$

It has been proven that the AIC is an asymptotically efficient selection method when the underlying process is an infinite order process. Nevertheless AIC has shown that it does not guarantee a consistent selection as the sample size diverges when the models in equations 5.2 and 5.3 are nested.

The BIC criteria (Schwartz (1978)) has been introduced to avoid the overestimation of the AIC criteria. A different penalty term impacts the likelihood in equation 5.4, then:

$$BIC_n = ln\left(\hat{\sigma}_n^2\right) + c\frac{nln(T)}{T}$$

with $c \geq 1$.

The HQ criteria, (Hannan and Quinn, 1982) proposes another penalization:

$$HQ_n = ln\left(\hat{\sigma}_n^2\right) + c\frac{ln\left(ln(T)\right)}{T}$$

When the competing models in equations 5.2 and 5.3 are non-nested Linhart (1988) and Vuong (1989) propose a general testing procedure (called KLIC) based on goodness-of-fit penalized likelihood function (denoted IC) in terms of minimization of a loss function, then we get:

$$IC_n = \sum_{t=1}^{n} q_{1t}(\beta_{1n}^*) - q_{2t}(\beta_{2n}^*) + c_n \tag{5.9}$$

where $q_t(\beta)$ is equal to minus the likelihood function for each modelling. We expect $IC < 0$ if we want to retain the model M_1.

The penalty term in all these criteria favours the selection of a parsimonious model, and it is typically a non-stochastic sequence: $c_n = (p-q)$ for AIC, $c_n = ((p-q)log(n))/2$ for BIC and $c_n = (p-q)clog(log(n))$ with $c > 1$ for HQ. These criteria require that the competing models are completely parametrized. The innovations errors need to belong to some parametric family of distributions. These criteria cannot be used when the models are estimated using moment conditions, GMM for instance. In certain cases it could be useful to have a model selection testing frame work that allows for a wide variety of estimation techniques. When we work with qualitative variables, the likelihood is not always possible adjusted, and thus these kind of criteria cannot be used. Thus, these criteria can be used for GDP modelling only under assumptions of independence and identically distributed for the data.

5.4.3 Encompassing tests using a predictive approach

There are several forecast-encompassing tests that have been developed to determine whether one model produces statistically superior forecasts relative to another. Such tests have been proposed by Diebold and Mariano (1995), West (1996), West (2001a), West (2001b), Harvey et al. (1998) and Harvey and Newbold (2000). A crucial requirement of these tests is that the models being compared be non-nested with the Weiss (2001) approach, and nested with the McCracken (2004) approach.

The forecast encompassing test of West (2001a,b) works as follows. We consider the following two models:

$$M_1 \ Y_t = Z'_{1t}\beta_1 + \epsilon_{1t} \tag{5.10}$$

$$M_2 \ Y_t = Z'_{2t}\beta_2 + \epsilon_{2t} \tag{5.11}$$

where for $i = 1, 2$, β_i and Z_{it} collect the parameters and the explanatory variables of model i. In addition, ϵ_{it} are white noise processes, for instance, model M_1 corresponds to an AR(p) for GDP and model 2 to a regression of the GDP on oil. We estimate the parameters once using the data until time T and using the n remaining data for out-of-sample forecast evaluation. Therefore, for t on $(T + 1)$ to $(T + n)$, we have the forecast $f_{it} = Z'_{it}\hat{\beta}_i$ for $i = 1, 2$. Moreover, we can obtain the h-step-ahead forecast f_{it}^h using direct or recursive forecast and denote by e_{it} the corresponding forecast error. The combined forecast is defined by $Y_t = (1 - \alpha)f_{1t}^h + \alpha f_{2t}^h + w_t$ which is equivalent to $e_{1t} = \alpha(e_{1t} - e_{2t}) + w_t$ with wt an error process and e_{1t} a process with zero conditional and unconditional means. Model 1 forecast encompasses model 2 if $\alpha = 0$ that is if $Ed_t = 0$ with $d_t = e_{1t}(e_{1t} - e_{2t})$.

Clark and McCracken (2001) and McCracken (2004) have developed another approach that can, respectively, test for encompassing forecasts between nested models, and test for equality of mean squared errors for nested models. For McCraken test we denote d_{t+k} the difference between the squared forecast errors at the $t + k$ of the base case model (i.e. model without financial variables, for instance, referring to GDP modelling), and the alternative model (model augmented with one or two variables):

$$d_{t+k} = \hat{\varepsilon}_{1,t+k}^2 - \hat{\varepsilon}_{2,t+k}^2. \tag{5.12}$$

With n forecast periods, the statistics for testing the equality of mean squared errors (MSEs) between the base case and the alternative model is computed as

$$MSE = n \sum \frac{n^{-1}\sum_{t=R-k}^{T}(\hat{\varepsilon}_{1,t+k}^2 - \hat{\varepsilon}_{2,t+k}^2)}{n^{-1}\sum_{t=R-k}^{T}\hat{\varepsilon}_{2,t+k}^2}. \tag{5.13}$$

where R represents the first out-of-sample forecast period. Intuitively note that the numerator represents the difference in MSEs between the base case and alternative model, and the denominator represents the MSE of the alternative. If both models produce equally accurate forecasts, then the numerator and test-statistic are zero; if the base case model has a lower MSE, then the statistic will be negative, and it will be positive if the alternative has a lower MSE. The distribution is not standard due to the fact that the models are nested, and so we can use the critical values computed by McCracken (2004).

5.4.4 Model selection for linear and non-linear dynamics with non-parametric techniques

We first introduce the Vuong (1989) approach, and then we detail the generalisation of the testing procedure for non-parametric modellings.

A general encompassing testing procedure has been developed based on a likelihood-ratio statistic to test the null that competing models are equally close to the DGP, in the sense of having the same Kullback-Leibler distance from the true DGP. If the null is true the two models are equivalent. This approach permits to compare models for dependent and heterogeneous processes, for linear and non-linear models, correctly specified and misspecified. For nested and overlapping models, the statistic is proposed by Marcellino and Rossi (2008), and for non-nested models by Rivers and Vuong (2002). We consider a stationary data set $(Y1, \ldots, Yn)$, and two parametric models with parameters β_j, $j = 1, 2$ and their estimates β_j^* based on the loss functions $Q_{jn}(\beta_j) = n^{-1} \sum_{t=1}^{n} q_{jt}(\beta_j)$, $j = 1, 2$, where q_{jt} is the minus of the log-likelihood function of each modelling. The objective is to select the model that minimizes the probability limit of the loss function $Q_{jn}(\beta_j) = E[n^{-1} \sum_{t=1}^{n} q_{jt}(\beta_j)] = \bar{Q}_{nj}(\beta_j)$, and the Vuong test

$$SC(\beta_1^*, \beta_2^*) = E[Q_{1n}(\beta_1^*) - Q_{2n}(\beta_2^*)] = 0 \tag{5.14}$$

means that the two models are equally good. The limit distribution of equation 5.8 under H_0 and H_1 is provided in Vuong (2001).

In the previous approach we focus on the encompassing test for testing one parametric model against another parametric model. Now, we are interested in the testing of two non-parametric models or "crossed" situations, where a parametric model is tested against a non-parametric one. The encompassing test statistics for these cases are based on appropriately normalized difference between an estimator of M_2 parameters (eventually functional), and its pseudo-true value under M_1. The specification tests for non-parametrically estimated models have meaning only when the smoothing parameter is not arbitrarily chosen. We assume that we have two models M_1 and M_2 with respective data densities $f(y, \beta)$ and $g(y, \beta_1)$. We denote $\hat{\beta}_1$ the statistic of interest obtained within the context of M_2 (an estimate of β_1 for instance), and let $\beta_1(\beta)$ be its pseudo-true value on M_1. For the purpose of analysis, the latter is defined as the plim of $\hat{\beta}_1$ on M_1. We consider the difference between $\hat{\beta}_1$ and $\beta_1(\hat{\beta})$. We say that M_1 (parametrically) encompasses M_2 with respect to $\hat{\beta}_1$ if that difference is not significant relative to its asymptotic sampling distribution on M_1.

We recall the base result for parametric modellings. Given the random variables Y_i, $i = 1, \ldots, n$, that M_1 and M_2 are two linear models characterizing Y_i such that Y_i is respectively based on by $X = (X_i, i = 1, \ldots, p)$, and $Z = (Z_i, i = 1, \ldots, q)$. Then, for the statistic

$$\hat{\delta}_{\beta, \beta_1(\beta)} = \hat{\beta}_1 - \hat{\beta}_1(\hat{\beta})$$

where $\hat{\beta}_1$ is an estimate of β_1 and $\hat{\beta}_1(\hat{\beta})$ an estimate of its pseudo value, we have the following result, under regular conditions:

$$\sqrt{n}\hat{\delta}_{\beta,\beta_1(\beta)} \to N(0, \sigma^2\Omega)$$

and

$$\frac{n}{\sigma^2}(\hat{\delta}'_{\beta,\beta_1(\beta)}\omega + \hat{\delta}_{\beta,\beta_1(\beta)} \to \chi_l^2,$$

Where $\Omega = var(\mathbf{Z})^{-1}E[var(\mathbf{Z}|\mathbf{X})]var(\mathbf{Z})^{-1}$, Ω denotes an estimate of Ω, Ω^+ denotes a generalized inverse, l is the rank of Ω, σ^2 is the variance of Y_t.

Now, if M_1 and M_2 denote two non-parametric regression models based on $\mathbf{X}$, and $\mathbf{Z}$ respectively and considering regularity conditions for the estimates of the densities, then the distribution for the statistic: $\hat{\delta}_{f,g}(Z) = \hat{g} - \hat{G}(\hat{f})(\mathbf{Z})$ where $\hat{f}(\mathbf{X})$ and $\hat{g}(\mathbf{Z})$ can be estimated using kernels (Bontemps et al. (2008)) or nearest neighbors methods (Guégan and Rakotomarolahy (2010a)).

If model M_1 is chosen to be linear and M_2 is non parametrically specified, we test that a linear model on the $\mathbf{X}$'s space explains any result of a model based on the $\mathbf{Z}$'s space, even if the latter is very general. In that case, the encompassing statistic is again based on the difference between two non-parametric estimators $\hat{\delta}_{\hat{f},g} = \hat{g}(\mathbf{Z}) - \hat{G}(\hat{\beta})(\mathbf{Z})$.

If we consider the case of a non-parametric model for M_1 and a parametric one for M_2 then, the model M_1 is not constrained while its competitor is linear. This is, a priori, the most favourable situation for M_1 to encompass M_2. The test statistic is parametric by construction as M_2 is parametric, and like in the other cases the encompassing statistic is based on the difference between an estimator of β_1 and the pseudo-true value: $\hat{\delta}_{f,g} = \hat{\beta}_1 - \hat{\beta}_1(\hat{f})$.

5.5 Discussion around the construction of composite indicators with the model and variable selection approaches

The choice of a model and its adequacy for a data set is determinant for several purposes and in particular for forecasting. In this section we propose an approach based on a testing methodology through the encompassing tests, providing a new way to determine economic indicators for GDP modelling and forecasting.

We concentrate on two objectives: the models election considering non nested modellings using the encompassing test, and the variable selection problem. For this exercise we retain five variables which are well known to affect the behavior of the GDP (Guégan (2011)). In the basket of models we retain two linear parametric modellings (AR and VAR) and one non-parametric modelling, the k-NN approach which is interesting because known to take non-linearity into account and for which the asymptotic distribution of the encompassing test has been obtained (Guégan and Rakotomarolahy (2010b)).

5.5.1 The data

We use quarterly data for GDP from 1995Q1 to 2010Q1 and monthly data from 1995M1 to 2010M3 for the other previous cited variables. We consider CAC40 and DAX as proxy for Euro area market index by taking their mean and Brent oil price index as proxy for the Euro area oil prices. Since the oil and housing price indexes are in dollar we use the monthly Euro/Dollar exchange rate to transform these prices in Euro currency. All along the exercise we update the relationship between the variable of interest (GDP) and the other variables (financial variables, housing prices and oil prices), by screening the variables that have a strong relationship with GDP. The data cover different economic contractions or cycles caused from various

shocks in other sectors, in particular the two remarkable crisis, the dotcom crisis in the last decade and the recent subprime crisis.[1] Their descriptive statistics are provided in Tables 5.A.1 and 5.A.10.

Before modelling the data sets, several properties of the data are tested, in particular a possible linear or non-linear relationship between the GDP and the exogenous variables and also the question of stationarity. To tackle these issues we transform the monthly variables into quarterly variables by aggregating the monthly values into a quarter. We also address issues on data transformation particularly by looking on the dynamics of log transform. In the following, Y_t will be the variable of interest, say the GDP and X_t will contain the variables to be tested: the interest rate spread (S), the oil price index (O), the housing price index (H), the stock market index (X) and the effective exchange rate (EER).

Concerning the linearity testing, there exist several linear tests: Ramsey (1969), White (1982), White (1990), Lee et al. (1993), Teräsvirta (2006) and Blake and Kapetanios (2003), among others. The Teräsvirta (2006) test often has been found more powerful than the other linearity tests, so we use it to test existence of linearity when we regress the GDP on economic and financial variables, considering the log transform and several forecast horizons. We report the corresponding p-value of the test in Table 5.A.2 for the Euro zone and in Table 5.A.11 for US zone : a value bigger than 0.05 means that we accept linearity.

For the Euro zone, the results reveal presence of both linear and non-linear relationships and we accept the linearity modelling for the spread at lags 1 and EER variables except at lag 8. We always reject the null for the housing and oil variables. For the US zone, the results reveal presence of both linear and nonlinear relationships as we accept the linearity modelling for the spread, index and EER variables. We always reject the null for the oil and housing variables. Thus, in the following modelling, we will consider both linear and nonlinear relationships.

Before modelling these data, we analyse the stationarity of the data sets using the Augmented Dickey-Fuller (ADF) test testing the non-stationarity under the null, and stationarity under the alternative. Table 5.A.3 provides the results of this test on logarithm and logarithm difference for the variables of the Euro zone and Table 5.A.12 for US zone. From the Table 5.A.3, for a given risk level $\alpha = 0.05$, we accept the null meaning presence of non-stationary for the GDP. For the same risk level, we reject the null in favor of the alternative stationary condition for the oil, and we accept existence of non-stationarity for the other economic and financial variables. These results justify modelling the GDP using a trend, and when we use the stochastic transformation for GDP we will also consider a stochastic transformation on the exogenous variables.

The results of the ADF test are provided in Table 5.A.3 for euro zone and in Table 5.A.12 for US zone. For a given risk level $\alpha = 0.05$, we accept the null for variables in the logarithm, i.e. the presence of non-stationary. In contrast, for the same risk level, we reject the null in favor of the alternative stationary condition for variables in growth.

5.5.2 Encompassing test and linear modellings

For calibration purposes, we divide the raw sample into two subsamples: the first subsample from 1995Q1 to 2005Q1 will be used for model building and the second one from 2005Q2 to 2010Q1 for predictive evaluation test. Using the first sample we estimate linear AR(p) models for the GDP with or without exogenous variables at an horizon h varying from 1 to 8. The parameters are obtained using least squares estimation method.

---

[1]The data are collected from Eurostat base and Datastream for financial data. The quarterly US real GDP at constant price and seasonally adjusted come from Datastream (code: GDP CONA); the monthly US three month T-Bill rate come from Datastream (code: FRTBS3M); the monthly ten year US government zero coupon bond come from Datastream (code: USBD10Y); the monthly S and P 500 index come from the Ecowin base; the monthly WTI oil spot prices index come from Datastream (code: OILWTXI); the monthly United State-DS Real Estate index come from Datastream (code: RLES-TUS); the monthly real effective exchange rates come from the Bank for International Settlements (http://www.bis.org/ and M:R:N:US)

Selection of linear modelling for log GDP

We begin to test linear models and we use several criteria for model selection. We retain the out-of-sample selection FPE, and the criteria BIC and AIC_c (this last one is more appropriate than AIC for small samples, Hurvich and Tsai (1989)). We consider also the unrestricted fixed order four that we often meet in autoregressive modelling of the quarterly GDP, Marcellino and Rossi (2008). For the Euro zone, Table 5.A.4 summarizes the selected order for the AR modellings for different horizons h, and Table 5.A.13 for the US zone.

From Tables 5.A.4 and 5.A.13, we observe that the information criteria AIC_c and BIC select the same lag order. In contrast, the out-of-sample FPE criterion can yield higher order for smaller horizons. Then, we analyse the validity of various models associated to various selection criteria. So, we carry out a specification tests to verify the whiteness of the residuals and existence or not of ARCH effects test. The results of the Ljung-Box test and the McLeod-Li test are provided in Table 5.A.5 for the Euro zone and in Table 5.A.14 for the US zone. In both cases, it appears that we can retain AR(p) models with a maximum value of 4 for p. In the following we focus on the log GDP. The results are similar with the log difference.

For the Euro zone we validate two models, an AR(1) and an AR(4) fitted on the centered logarithm GDP, $(Y_t)_t$. After eliminating the insignificant parameters, we retain the following models for $(Y_t)_t$, for $h = 1$:

$$
\begin{aligned}
M_1 : Y_t = \quad & \underset{(0.00000)}{0.0001} \ t \ + \ \underset{(0.008)}{0.96} \ Y_{t-1} + \epsilon_{1t} \\
M_2 : Y_t = \quad & \underset{(0.000)}{0.0001} \ t \ + \ \underset{(0.17)}{1.11} \ Y_{t-4} + \epsilon_{2t}
\end{aligned}
\tag{5.1}
$$

The values in brackets correspond to the standard errors of the estimated coefficients and (it)t, i =1, 2, 3 are white noise processes. These two models are non-nested as neither one is a special case of others.

For the US zone we retain three model specifications. Using the same notations as before, we get:

$$
\begin{aligned}
M_1 : Y_t = \quad & \underset{(0.00003)}{0.00025} \ t \ + \ \underset{(0.006)}{0.95} \ Y_{t-1} + \epsilon_{1t} \\
M_2 : Y_t = \quad & \underset{(0.00006)}{0.0002} \ t \ + \ \underset{(0.16)}{0.97} \ Y_{t-3} + \epsilon_{2t} \\
M_3 : Y_t = \quad & \underset{(0.00006)}{0.00019} \ t \ + \ \underset{(0.18)}{0.98} \ Y_{t-4} + \epsilon_{3t}
\end{aligned}
\tag{5.2}
$$

The three models are non-nested as neither one is a special case of others.

Encompassing tests for non-nested linear models

Before applying the encompassing test, we determine what is the model which exhibits the smaller error variance. The standard errors for EU models M_i, $i = 1, 2$, are respectively $\sigma_1 = 0.0050$ and $\sigma_2 = 0.0053$. Then, M_1 corresponds to the best fit. Before applying the encompassing test on the different AR modellings, we regress The GDP on the lag one and four. We get

$$
Y_t = 1.28.10^{-4}t + 1.041Y_{t-1} - 0.075Y_{t-4} + \varepsilon_t.
\tag{5.3}
$$

The coefficient $\phi_1 = 1.041$ is significant but not the coefficient $\phi_4 = 0.075$. Thus the model AR(1) encompasses the model AR(4). To confirm this evidence we apply the encompassing test on linear models. The results are provided in Table 5.A.6. From the results, we accept the null $M_1 \mathcal{E} M_2$ that is, M_1 -the AR(1) model-

encompasses M_2 -the AR(4) model. Indeed, in that case the p-value is greater than 0.05. In contrast, we reject $M_2 \mathcal{E} M_1$: the AR(4) model does not encompass the AR(1) model. Thus, in the following we retain the AR(1) model for the EU GDP.

We do the same exercise for the US zone. The standard errors for models M_i, $i = 1, 2, 3$ are $\sigma_1 = 0.0050$, $\sigma_2 = 0.0047$ and $\sigma_3 = 0.0048$, respectively. Then, among the three models, M_1 has the worst fit. In contrast, M_2 has the best fit. For the encompassing test, we present the results in Table 5.A.15. From this table we accept the null $M_2 \mathcal{E} M_3$ that is, M_2 encompasses M_3. In contrast, we reject $M_3 \mathcal{E} M_2$ and $M_1 \mathcal{E} M_2$ and therefore M_3 encompasses neither M_1 nor M_2. Thus in the following we retain the AR(3) model for the US GDP.

This first analysis permits to show that for the Euro zone the AR(1) dominates all the univariate modellings, and the AR(3) dominates for the US zone. Nevertheless the predictive performance of the AR models is poor, thus we now consider multivariate modellings.

VAR modelling with exogenous variables

At this step inside the VAR modelling we include exogenous variables which are economic and financial variables. We use stepwise feed forward selection and backward elimination methods for selecting the lag order of the exogenous variables.

For the Euro zone, we retain a VAR(4) using the oil, a VAR(1) with the housing, a VAR(2) respectively with the index and EER variables; and for the US zone we retain a VAR(1) whatever the variables we consider. With this approach we are only interested to test the predictability of the modellings which will be the criteria retained to choose the best VAR modelling. For this out-of-sample forecast evaluation, we compute the RMSE on the subsample 2005-2010. Table 5.A.7 summarizes the results for the Euro zone and Table 5.A.16 for the US zone.

From Table 5.A.7, when we compare the results obtained with the AR(1) on GDP with the VAR modellings the best fits are obtained with the VAR modellings. For horizon one the best fit is obtained using spread or oil, for horizon 2 with oil or housing, and for horizons 4 and 8 using oil. Thus for predictive ability the VAR modelling including oil as exogenous variable provides the best forecasts for short, medium and long term.

From Table 5.A.16, we observe that we get some improvement for the forecast of US GDP using any variable except the market index. We have the best forecast with oil whatever the horizons we consider. Housing is also an interesting variable in the short term for predictive ability. Spread is not so determinant. Thus the impact of economic variables dominates on the others.

k-NN modelling

We now consider the non-parametric k-nearest neighbor regression method. Following the presentation of the previous Sections, we use the following nonlinear representation for (Y_i, X_i):

$$Y_t + 1 = \beta' Z_t + m_Y(Z_t) + u_t \text{ and } X_{t+1} = m_X(X_t) + v_t \tag{5.4}$$

where $Z_t = (Y_t, X_t)$ with Y_t the log GDP, and X_t the logarithm of financial/economic variables, $(u_t)_t$ and $(v_t)_t$ are error processes and the unknown parameters are β. We use the same information set for the estimation sample and the out-of-sample as before. For estimation we estimate the parameters in two steps. First we regress Y_{t+1} on Z_t and estimate the parameter β, then using the residuals we estimate the functions m_Y and m_X with the nearest neighbor method with $k = 3$ working in dimension 2 $(d = 2)$. As soon as the estimation step is done we make forecasts. For the out-of-sample evaluation, we compute the corresponding RMSE. For the Euro zone the results are given in Table 5.A.8, and in Table 5.A.17 for the US zone. We have considered recursive forecasts when the horizon h is greater than 1.

Looking at the RMSE obtained with the k-NN method applied to the Euro GDP alone or when we have two variables, the forecast accuracy increases in all cases when we add an exogenous variable. The best forecasts are obtained with the pair (GDP, spread) for all horizons. We observe also that the results are very close each other. For the US economy, we obtain also better forecasts for GDP with exogenous variables. The best forecasts are obtained with housing at horizon $h = 1$, with EER at horizon $h = 2, 8$, and with spread and index at horizon $h = 4$.

In this exercise the best forecast changes with the forecast horizon and therefore we can think that most of the exogenous variables are useful for forecasting GDP. With this exercise, we observe that the k-NN modelling improves a lot the accuracy of the forecasts obtained with linear modellings and that the influence of certain financial or economic variables is not negligible in the performance we observed.

Encompassing test for parametric modelling against k-NN modelling

We found from the previous results that nearest neighbor method improves the accuracy of the GDP forecast when we compare it with the values obtained with linear modellings for the same horizons. We know also that the encompassing model ought to be able to explain the predictions of the encompassed model, thus we are interested on testing if the nearest neighbor regression technique can encompass linear modellings.

We apply the encompassing test, developed in the previous Section, by testing the null that is the nearest neighbor regression with exogenous variables encompassing the linear AR(p) for the GDP. Using previous notations, under this null, we consider the statistic

$$S = \frac{\hat{\delta}}{\sqrt{\hat{\Omega}}} = \sum_t (Z_t Z_t)^{-1} \sum_t \hat{u}_t Z_t / \sqrt{\hat{\Omega}} \tag{5.5}$$

which follows approximately a standard normal distribution, with variance $\hat{\Omega}$, where

$$\hat{\Omega} = \hat{\sigma}^2 (\sum_t (Z_t Z_t)^{-1} \tag{5.6}$$

and $Z_t = Y_{t-p}$ is the regressor of the model AR(p), $\hat{u}_t$ is the estimated residuals from nearest neighbor model and $\hat{\sigma}^2$ is a $k - NN$ estimate of the conditional variance $\sigma^2 = var(Y_t | Y_{t-1}, ..., Y_{t-p}, X_{t-1})$. The results of this test are reported in Table 5.A.9 for the Euro zone and in Table 5.A.18 for the US economy.

Recalling that $X \mathcal{E} AR(p)$ is accepted as soon as the p-value is greater than 0.05 we conclude that for the Euro zone the AR(1) model encompasses the $k - NN$ modelling for the short term and the $k - NN$ modelling encompasses the AR(1) model for long term, h =8. Thus, for short and medium terms, it appears that all the information contains in $k - NN$ is included in the AR(1), and the contrary for the long term.

For the US data, in all cases we accept the null. This implies that the k-NN modelling of GDP with exogenous variables encompasses the model AR(3) fitted for GDP. In other words, information contained in the model AR(3) is already included in the $k - NN$ modelling of GDP with financial/economic variables.

Thus, for short term modelling inside the Euro zone the linear modelling dominates the non linear modelling and this is never true at long term. The situation appears different for the US economy because in that latter case the non linear modelling dominates the linear modelling for all horizons. With this exercise we also note that working with model selection leads to prefer the linear modelling for the Eurozone, but if we have in mind the variables selection with a predictive approach we will prefer the $k - NN$ modelling. This conclusion can be contradictory but from a statistical point of view, we can conclude that a linear model is preferable (by testing) for structural reasons, and a non linear approach is useful for predictive purposes. We do not encounter this situation with the US variables.

5.6 Conclusion

We provide some keys to build composite indicators which affect the behavior of the GDP. With this work, we have focused on two points: to introduce a procedure for model selection for non-nested models, parametric or non-parametric modellings for GDP. We have also discussed the problem of variables selection considering both approaches, one with a criteria of predictability and the other one using encompassing tests.

We have applied these methodologies on two data sets and we highlight the power of encompassing tests to determine the model which dominates the other one in a basket of models for a given set. We also retain that it can arrive at some opposite conclusion with respect to the methodologies used.

Concerning the predictability of the GDP we conclude that the use of certain economic and financial variables can improve the short or the long term with respect to the data used and also to the modelling. Thus, economic and financial variables contain information unexplained in the GDP dynamics which can improve its accuracy forecast, and we can use them to determine composite indicators for GDP forecasting.

The dominance of the $k - NN$ method for forecasting the GDP for US data and EU data on long term proves that non-linearity exists and has to be taken into account inside this kind of data sets. The use of encompassing test has confirmed this fact for the US data.

This work has opened a lot of new routes which need to be developed. The extension of the encompassing test for other non parametric modellings, like the radial basis function and the neural network method which are known to improve the predictability of the GDP. Other technical methods like pooling methods can also be considered, nevertheless all these techniques are time consuming and for the moment their applicability is restricted to linear modellings, and their use is still limited.

Annex

5.A Tables

Table 5.A.1: Euro zone: Summary statistics of the economic and financial variables.

Variable	Min	Max	Mean	Std Dev	Skewness	Kurtosis
GDP	7.26	7.57	7.43	0.09	-0.001	0.0016
Spread	-0.38	3.40	1.36	0.90	-0.44	17.95
Oil	12.25	4.35	3.31	0.54	-0.09	1.95
Housing	5.44	7.01	6.00	0.44	0.43	1.01
Index	7.55	8.83	8.33	0.34	-0.15	0.40
EER	4.35	4.75	4.57	0.10	60.002	0.002

Table 5.A.2: Euro zone: Linear Test when we regress the GDP on other variables. P-value of test for horizon h=1,2,4,8

	1	2	4	8
Spread	0.057	0.006	0.00	0.00
Oil	0.001	0.02	0.09	0.07
Housing	0.00	0.00	0.00	0.00
Index	0.01	0.001	0.00	0.00

Table 5.A.3: Euro zone: Augmented Dickey-Fuller unit root test for the five variables

Variable	Variable in logarithm		Variable in growth	
	statistic	p-value	statistic	p-value
GDP	-0.89	0.94	-3.28	0.08
Oil	-3.86	0.02	-3.91	0.01
Housing	-1.63	0.72	-3.02	0.16
Index	-2.59	0.33	-2.92	0.20
EER	-2.42	0.40	-3.21	0.09

Table 5.A.4: Euro zone: Lags selection for AR(p) modeling of the logarithm GDP. Lags retained for AR(p) models

Model	Selection criteria			
h	AICc	BIC	FPE	lag.max
1	1	1	1	4

Table 5.A.5: Euro zone: Specification test for the logarithm GDP model. Notes: Values of the test statistics (df=10) with the corresponding p-values in brackets.

Model	Error autocorrelation test Ljung-Box's test			No ARCH effect test McLeod-Li's test		
	IC	FPE	lag.max	IC	FPE	lag.max
1	6.51	6.51	3.63	0.64	0.64	0.71
	(0.77)	(0.77)	(0.96)	(1)	(1)	(1)

Table 5.A.6: Euro zone: Encompassing test for the logarithm GDP model. Encompassing test with their p-values in brackets.

	$M_2 E M_1$	$M_1 E M_2$
t_{encomp}	11.69 (10-12)	-0.93 (0.35)

Table 5.A.7: Euro zone: VAR modelling out-of-sample RMSE between 2005Q2 and 2010Q1. Notes: For GDP, RMSEs are from univariate modelling.

VAR model	RMSE for horizon h			
Z_t	1	2	4	8
GDP-AR(1)	0.028	0.039	0.052	0.067
(GDP,Spread)	0.010	0.018	0.036	0.064
(GDP,Oil)	0.008	0.015	0.028	0.040
(GDP,Housing)	0.008	0.015	0.062	0.096
(GDP,Index)	0.012	0.025	0.042	0.055
(GDP,EER)	0.014	0.022	0.027	0.029

Table 5.A.8: Euro zone: Nearest neighbor out-of-sample RMSE between 2005Q2 and 2010Q1.

kNN model	RMSE for horizon h			
	1	2	4	8
GDP-kNN	0.0163	0.021	0.032	0.047
(GDP,Spread)	0.006	0.010	0.013	0.020
(GDP,Oil)	0.008	0.014	0.018	0.028
(GDP,Housing)	0.008	0.014	0.019	0.028
(GDP,Index)	0.008	0.014	0.018	0.030
(GDP,EER)	0.008	0.013	0.018	0.030

Table 5.A.9: Euro zone: Encompassing test nearest neighbor vs auto-regressive.

test	Statistic for horizon h (p-value in parenthesis)			
kNNEAR(1)	1	2	4	8
SpreadEAR(1)	-2.09 (0.05)	-2.42 (0.025)	-1.69 (0.10)	-1.12 (0.27)
OilEAR(1)	-2.05 (0.05)	-2.77 (0.01)	-2.48 (0.02)	-0.62 (0.53)
HousingEAR(1)	-2.28 (0.034)	-2.73 (0.013)	-2.51 (0.02)	-0.01 (0.99)
IndexEAR(1)	-2.51 (0.02)	-3.17 (0.004)	-2.60 (0.017)	-0.01 (0.99)
EEREAR(1)	-1.69 (0.107)	-2.52 (0.02)	-2.33 (0.03)	-0.48 (0.63)

Table 5.A.10: US zone: Summary statistics of the economic and financial variables.

Variable	Min	Max	Mean	StdDev	Skewness	Kurtosis
GDP	9025.30	13415.30	11551.23	1378.79	-0.32	1.81
Spread	0.25	4.33	2.24	1.13	0.23	1.81
Oil	12.94	123.95	41.02	26.16	1.23	3.98
Housing	437.43	1710.40	973.33	291.19	0.29	2.81
Index	486.17	1505.45	1092.38	264.89	-0.48	2.37
EER	89.77	119.09	102.99	7.54	0.31	2.19

Table 5.A.11: US zone: Linearity test in mean of the regression of GDP on five variables.

Model	p-value of test for horizon h			
X_t	1	2	4	8
Spread	0.18	0.06	0.00	0.00
Oil	0.00	0.00	0.00	0.00
Housing	0.00	0.00	0.00	0.00
Index	0.21	0.23	0.25	0.00
EER	0.24	0.25	0.24	0.03

Table 5.A.12: US zone: Augmented Dickey-Fuller unit root test for the five variables.

	Variable in logarithm		Variable in growth	
Variable	statistic	p-value	statistic	p-value
GDP	-0.90	0.94	-3.72	0.03
Oil	-3.23	0.09	-4.56	0.01
Housing	-2.78	0.26	-3.33	0.07
Index	-2.52	0.36	-3.14	0.11
EER	-2.23	0.47	-4.40	0.01

Table 5.A.13: US zone: Lags selection for AR(p) modelling of the logarithm GDP. Lags for the AR(p) model

Model	Selection criteria			
i	AICc	BIC	FPE	lag.max
1	1	1	3	4
2	1	1	3	4
4	1	1	1	4
8	1	1	1	4

Table 5.A.14: US zone: Specification test for the logarithm GDP model. Notes: These values are the statistics of the test (df=10) with the corresponding p-values in parenthesis.

Model	Error autocorrelation test Ljung-Box's test			No ARCH effect test McLeod-Li's test		
	IC	FPE	lag.max	IC	FPE	lag.max
1	11.64	7.11	7.49	8.46	4.56	4.83

Table 5.A.15: US zone: Encompassing test for the logarithm GDP model. Notes: Statistics of the test with their p-values in brackets.

	M_1EM_3	M_3EM_1	M_3EM_2	M_2EM_3
t_{encomp}	0.79 (0.43)	11.50 (0.00)	3.95 (0.00)	1.03 (0.31)

Table 5.A.16: US zone: VAR modeling out-of-sample RMSE between 2005Q2 and 2010Q1. Notes: For GDP, RMSEs are from univariate modelling.

VAR model	RMSE for horizon h			
Z_t	1	2	4	8
GDP-AR(3)	0.0187	0.0187	0.0289	0.0385
(GDP,Spread)	0.0155	0.0304	0.0595	0.1302
(GDP,Oil)	0.0082	0.0126	0.0198	0.0285
(GDP,Housing)	0.0094	0.0172	0.0510	0.0710
(GDP,Index)	0.0276	0.0425	0.0741	0.1020
(GDP,EER)	0.0097	0.0180	0.0239	0.0285

Table 5.A.17: US zone: Nearest neighbor out-of-sample RMSE between 2005Q2 and 2010Q1.

kNN model	RMSE for horizon h			
	1	2	4	8
GDP-kNN	0.0130	0.0176	0.0276	0.0479
(GDP,Spread)	0.0077	0.0132	0.0144	0.0232
(GDP,Oil)	0.0081	0.0134	0.0163	0.0230
(GDP,Housing)	0.0074	0.0127	0.0147	0.0198
(GDP,Index)	0.0075	0.0118	0.0144	0.0219
(GDP,EER)	0.0075	0.0117	0.0147	0.0211

Table 5.A.18: US zone: Encompassing test nearest neighbor vs autoregressive.

test	Statistic for horizon h (p-value in parenthesis)			
kNNEAR(3)	1	2	4	8
SpreadEAR(3)	-0.75 (0.45)	-1.32 (0.20)	-1.83 (0.08)	-0.55 (0.58)
OilEAR(3)	-0.67 (0.50)	-1.00 (0.32)	-0.91 (0.37)	-0.22 (0.22)
HousingEAR(3)	-0.45 (0.65)	-0.91 (0.37)	-0.47 (0.63)	0.04 (0.96)
IndexEAR(3)	-0.36 (0.71)	-0.72 (0.47)	-0.59 (0.56)	0.01 (0.99)
EEREAR(3)	-0.48 (0.63)	-0.88 (0.38)	-0.56 (0.57)	-0.06 (0.94)

Bibliography

Akaike H. (1969), Fitting autoregressive models for prediction, Annals of the Institute of Statistical Mathematics, 21, 243 -247.

Akaike H. (1973), Information theory and an extension of the likelihood principle, in Petrov B.N. and Csaki F. (eds), Proceedings of the second symposium of information theory, Akademiai Kiado, Budapest, 257 281.

Akaike H. (1974), A new look at the statistical model identification, IEEE Transactions and Automatic Control, AC19, 716 -723.

Bickel P. J., Klaassen C.A.J., Ritov Y., Wellner J. A. (1992), Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore.

Billio M., Ferrara L., Guégan D., Mazzi G.L. (2009), Evaluation of Non-linear time-series models for real-time business cycle analysis of the Euro area, WP-2009-53, Publications du CES Paris 1 Panthéon-Sorbonne.

Blake A.P., Kapetanios G. (2003), A radial basis function artificial neural network test for neglected non-linearity, Econometrics Journal 6, 356-372.

Bontemps C., Florens J.P., Richard J.F. (2008), Parametric and non-parametric encompassing procedures, Oxford Bulletin of Economics and Statistics, 70 (Supplement) , 0305-9049.

Cai J. (1994), A Markov Model of Switching -Regime ARCH, Journal of Business and Economic Statistics, 12, 309 -316.

Clark T.E., McCracken M.W. (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, Journal of Econometrics 105, 85-110.

Diebold F.X., Mariano R.S. (1995), Comparing predictive accuracy, Journal of Business and economic Statistics, 12, 253 -263.

Durland J.M., McCurdy T.H. (1994), Duration-dependent transitions in a Markov model of U.S. GNP growth, Journal of Business and Economic Statistics, 12, 279 -288.

Engle R.F., Granger C.W.J., Rice J., Weiss A. (1996), Semi-parametric estimates of the relationship between weather and electricity sales, J.A.S.A., 81, 310 -320.

Ferrara L., Guégan D. (2004), Detection of the industrial business cycle using SETAR models, Journal of the business cycle measurement and analysis, 2, 353 -372.

Ferrara L., Guégan D., Rakotomarolahy P. (2010), GDP nowcasting with ragged-edge data: a semi-parametric modeling. Journal of Forecasting 29, 186-199.

Filardo A.J. (1994), Business-Cycle Phases and Their Transitional Dynamics, Journal of Business and Economics Statistics, 12, 299-308.

Girosi F, Anzelotti G. (1993), Rates of convergence for radial basis functions and neural networks Artificial Neural Networks for Speech and Vision 1993; 97-114.

Guégan D. (2003), Les chaos en finance: Approche statistique, Economica, Série Mathématique et Probabilité, 420 pages.

Guégan D. (2009), Multivariate Nearest Neighbors Method for Forecasting in Economics, Lecture at Banque de France, March 2009, Paris, France.

Guégan D., Rakotomarolahy P. (2010a), A Short Note on the Now-casting and the Forecasting of Euro-area GDP Using Non-Parametric Techniques. Economics Bulletin 30, No.1, 508-518.

Guégan D., Rakotomarolahy P. (2010b), Alternative methods for forecasting GDP, in "Nonlinear Modeling of Economic and Financial Time Series", eds. R. Barnett and F. Jawady, Series International Symposia in Economic Theory and Econometrics", Emerald publishers, Chapter 5.

Guégan D. (2011), Contagion Between the Financial Sphere and the Real Economy. Parametric and non Parametric Tools: A Comparison, Université Paris-1 Panthéon-Sorbonne (Post-Print and Working Papers) halshs-00185373, HAL.

Hannan E.J., Quinn B.G. (1979), The determination of the order of a regression, Journal of the Royal Statistical Society Series B, 41, 190–195.

Hamilton J.D. (1988), Rational Expectations Econometric Analysis of Change in Regime: An Investigation of the Term Structure of Interest Rates, Journal of Economic Dynamics and Control, 12, 385-423.

Hamilton J.D., Susmel R. (1994), Autoregressive Conditional Heteroscedasticity and Changes in Regimes, Journal of Econometrics, 64, 307 -353.

Härdle W., Müller M., Sperlich S., Werwatz A. (2004), Non-Parametric and Semi-Parametric Models. Springer Verlag, 2004.

Harvey D., Leybourne S.J., Newbold P. (1998), Tests for Forecast Encompassing, Journal of Business and Economic Statistics, Vol. 16, No. 2, 254-259.

Harvey D., Newbold P. (2000), Tests for Multiple Forecast Encompassing Journal of Applied Econometrics. VOL 15(PART 5), 471-482.

Hurvich C.M., Tsai C.L. (1989), Regression and time series model selection in small samples, Biometrika, 76, 297 -307.

Kim C.J., Nelson C.R. (1999), State space models with regime switching, the MIT Press, Cambridge.

Krolzig H.M., Toro J. (2000), A New Approach to the Analysis of Business Cycle Transitions in a Model of Output and Employment, Preprint, Nuffield College, Oxford University, U.K.

Lam, Pok-Sang (2004). A Markov Switching Model of GNP Growth with Duration Dependence, International Economic Review 45, 175-204.

Lee T., White H., Granger C.W.J. (1993), Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests, Journal of Econometrics 56, 269-290.

Linhart, H. (1988), A test whether two AIC's differ significantly. S. Afr. Stat. J. 22:153-161.

McCracken M. W. (2004), Asymptotics for Out of Sample Tests of Causality, manuscript, University of Missouri.

Marcellino M. , Rossi B. (2008), Model selection for nested and overlapping nonlinear, dynamic and possibly mis-specified models, Oxford Bulletin of economics and statistics, (Supplement) 70, 0305-9049.

Ramsey J.B. (1969), Test for Specification error in Classical Linear Least Squares Regression Analysis, Journal of the Royal Statistical Society, Series B. 31, 350-371.

Rivers D., Vuong Q. (2002), Model selection tests for non-linear dynamic models, Econometrics Journal, 5, 1–39.

Silverman B.W (1986), Density Estimation for Statistics and Data Analysis. Chapmann and Hall: London, 1986.

Schwartz G. (1978), Estimating the dimension of a model, Annals of statistics, 6, 461 -464.

Teräsvirta T. (2006), Forecasting economic variables with nonlinear models, in G. Elliott, C. W. J. Granger and A. Timmermann (eds), Handbook of Economic Forecasting, vol. 1, Elsevier, Amsterdam, 413–457.

Vuong Q. (1989), Likelihood ratio tests for model selection and non-nested hypothesis, Econometrica, 57, 307–333.

Wegman E.J., Wright I.W. (1983), Splines in statistics, J.A.S.A., 78, 351–365.

West K. D. (2001a), Tests of Forecast Encompassing When Forecasts Depend on Estimated Regression Parameters, Journal of Business and Economic Statistics 19, 29-33.

West K. D. (2001b), Encompassing tests when no model is encompassing., manuscript, University of Wisconsin.

West K.D. (1996), Asymptotic inference about predictive ability, Econometrica, 64, 1067 -1084.

White H. (1982), Maximum likelihood estimation of mis-specified models, Econometrica, 50, (1).

White H. (1990), A consistent model selection, in Grander C.W.J. (ed.), Modelling economic series, Clarendon Press, Oxford, 369 -383.

6

Variable Reduction and Variable Selection Techniques

Authors

George Kapetanios
King's College, University of London
george.kapetanios@kcl.ac.uk

Fotis Papailias
King's College, University of London and quantf research
f.papailias@kcl.ac.uk

Massimiliano Marcellino
Bocconi University, IGIER, and CEPR
massimiliano.marcellino@unibocconi.it

Table of Contents

6.1 Introduction

A large part of research inside academia, banks, central banks and elsewhere focuses on finding the appropriate set of predictors to be used in nowcasting or forecasting various key macroeconomic variables. This is also relevant for the construction of composite coincident indexes (CCIs), as the relevant predictors for a key variable such as GDP growth could be then also used to construct a CCI.

One approach could be a careful selection of a small group of predictors that could be modelled as a VAR. As an alternative, under the assumption of non cross-correlated idiosyncratic errors, small scale dynamic or static factor models for the pre-screened series could be used. Some examples can be found in Stock and Watson (1991), Camba-Mendez et al. (2001), Aruoba et al. (2009), Aruoba and Diebold (2010) and Camacho and Perez-Quiros (2010).

On the other hand, the seminal work by Stock and Watson (2002a) and Forni et al. (2005) suggests forecasting with a large set of predictors under the hypothesis that many variables might be driven by a reduced number of common factors or shared trends. These can be typically extracted from the data set using static or dynamic principal components.

The forecasting performance of other variable reduction methods suited to handle large datasets has been recently studied by Del Mol et al. (2006), Kapetanios, Marcellino and Papailias (2012a) among others. They include Bayesian shrinkage regression, partial least squares and others.

A comparison between small scale and large scale dynamic factor models can be found in Alvarez et al. (2012). Here we take a broader perspective and compare various variable reduction and selection methods, according to their forecasting performance in a relevant empirical application.

We start with a very large dataset of 195 monthly variables selected among the Principal European Economic Indicators (PEEIs), which represent the Euro Area (EA) economic activity (in the spirit of Stock and Watson (2002a) for the US).

Then, we categorise the monthly variables according to their nature, we create a small dataset of predictors where one representative variable within each group is selected (a similar methodology can be found in Alvarez et al. (2012)), and we re-evaluate the forecasting performance of all methods. We continue in this manner and we create four additional medium datasets using four estimates of the exponent of the cross-sectional dependence recently introduced by Bailey, Kapetanios and Pesaran (2012). In general, small datasets contain less than 10 indicators and medium datasets about 20-30.

The reduction methodologies we compare include Principal Components (PC), Partial Least Squares (PLS) and Bayesian Shrinkage Regression (BR). For variable selection, we choose the appropriate instruments by minimising the Akaike's (AIC), Bayesian (BIC) and Hannan-Quinn's (HQ) information criteria.

The presence of a large set of explanatory variables requires the use of non-standard optimisation algorithms, and we consider: Simulated Annealing (SA), Genetic Algorithm (GA), Sequential Testing (ST) in the sense of Hoover and Perez (1999) and the MC^3 algorithm. A detailed description of each algorithm is given in the following section. Recently, Buchen and Wolhrabe (2011) evaluated the boosting method in the Stock and Watson (2006) dataset and Kapetanios, Marcellino and Papailias (2012a) applied these methods in the Euro Area.

We apply the different selection and reduction procedures on the alternative large and smaller datasets in order to forecast: (a) the growth of seasonally adjusted quarterly GDP in the Euro Area 16 (EA-16, fixed composition), (b) the growth of seasonally adjusted quarterly consumption growth rate in the Euro Area 17 (EA-17, fixed composition), (c) the EA-16 industrial production growth relative to the previous period and (d) the monthly growth of the Harmonized CPI (HICP) in EA-16. Our main result is that often the best forecasts are obtained using the small and medium datasets. Smaller datasets highly improve the forecasting performance of variable selection models and slightly enhance the variable reduction models. Overall, the former appear

to perform better than the latter in our application. Composite Indicators (as defined in Chapter 2) can be constructed using the hereby reviewed methodologies.

The rest of the chapter is organized as follows. Section 2 presents the models and the computational settings used in our experiments. Section 3 contains the description of the forecasting evaluation, a brief description of the data, and a discussion of how to handle data unbalancedness. Section 4 discusses the forecast results and Section 5 summarizes our main findings and conclusions.

6.2 Methodology

This section describes the alternative methods we apply in the empirical analysis.

6.2.1 Variable Selection Methods

We consider the following regression model,

$$y_t = \alpha + \beta^{0'} x_t^0 + \epsilon_t, \quad t = 1, \ldots, T, \tag{6.1}$$

where x_t^0 is a k-dimensional vector of stationary predetermined variables. The superscript ^{0} denotes the true regression model. Let the set of all available variables at time t be represented by the N-dimensional vector $x_t = (x_{1,t}, \ldots x_{N,t})'$, where it is currently assumed that the set of variables in x_t^0 is also contained in x_t. The aim of the analysis is to determine x_t^0. Formally, let $\mathcal{I} = (\mathcal{I}_1, \ldots, \mathcal{I}_N)'$ denote a vector of zeros and ones (which we will refer to as string). Let $\mathcal{I}^0$ be the string for which $\mathcal{I}_i^0 = 1$, if $x_{i,t}$ is an element of x_t^0 and zero otherwise. We wish to estimate $\mathcal{I}^0$. Note that in small samples $\mathcal{I}^0$ may not represent the best fitting model for the data at hand.

To do this we consider the use of information criteria to select the variables that go in (6.1). The generic form of such criteria is usually,

$$IC(\mathcal{I}) = -2L(\mathcal{I}) + C_T(\mathcal{I}), \tag{6.2}$$

where $L(\mathcal{I})$ is the log-likelihood of the model associated with string $\mathcal{I}$ while $C_T(\mathcal{I})$ is the corresponding penalty term. The three most usual penalty terms are $2\tilde{m}(\mathcal{I})$, $ln(T)\tilde{m}(\mathcal{I})$ and $2ln(ln(T))\tilde{m}(\mathcal{I})$ associated with the Akaike (AIC), Bayesian (Schwarz (1978)) (BIC) and Hannan-Quinn (Hannan and Quinn (1979)) (HQ) information criteria. $\tilde{m}(\mathcal{I})$ is the number of parameters in the model associated with I, so that $\tilde{m}(\mathcal{I}) = \mathcal{I}'\mathcal{I}$. It is straightforward under relatively weak conditions on $x_{j,t}$ and $\epsilon_{j,t}$, and using the results of say, Sin and White (1996), to show that the string which minimises $IC(.)$ will converge to $\mathcal{I}^0$ with probability approaching one as $T \to \infty$ as long as (i) $C_T(\mathcal{I}) \to \infty$ and (ii) $C_T(\mathcal{I})/T \to 0$.

More specifically, the assumptions needed for the results of Sin and White (1996) to hold are mild and can be summarised as follows, assuming estimation of the models is undertaken in the context of Gaussian or pseudo maximum likelihood (which in the simplest case, of spherical errors, is equivalent to OLS):

  (i) Assumption A of Sin and White (1996) requires measurability, continuity and twice differentiability of the log-likelihood function and a standard identifiability assumption;

  (ii) A uniform weak law of large numbers for the log-likelihood of each observation and its second derivative;

  (iii) A central limit theorem for the first derivative of the log-likelihood of each observation.

(ii) and (iii) above can be obtained by assuming, e.g., that $x_{j,t}$ are weakly dependent, say, near epoch dependent, processes and $\epsilon_{j,t}$ are martingale difference processes. Hence, it is clear that consistency of model selection as long as the penalty related conditions hold is straightforwardly obtained. Note that unlike BIC and

HQ which consistently estimate the true model in the sense of Sin and White (1996), AIC is inconsistent, in this sense, since C_T remains bounded, as $T \to \infty$, contravening the first penalty related condition given in the preceding paragraph. Further, note that in most work dealing with variable selection and information criteria, stationarity of x_t^0 is usually assumed, as is in the preceding analysis, although the analysis may be extended to nonstationary variables.

The problem is of course how to minimise the information criterion. For small dimensional x_t, evaluating the information criterion for all strings may be feasible, as, e.g., in lag order selection. In the case of lag selection the problem is made easier by the fact that there exists a natural ordering of the variables, although in many cases such an ordering may not be the optimal basis for a search algorithm. In the general variable selection case, as soon as N exceeds say 50 or 60 units, this strategy is bound to fail. Since $\mathcal{I}$ is a binary sequence there exist 2^N strings to be evaluated. For example, when $N = 50$ and optimistically assuming that 100000 strings can be evaluated per second, we still need about 357 years for an evaluation of all strings. Clearly this is infeasible.

Although this is a minimisation problem, standard minimisation algorithms do not apply due to the discreteness of the domain over which the objective function (information criterion) needs to be optimised. To overcome this difficulty we use the following heuristic optimisation approaches: simulated annealing, genetic algorithm, the MC^3 and sequential testing. They are described in the following section.

6.2.2 Non-Standard Optimisation Algorithms

Since in our case N is very large, N=195, we should compute and compare 2^{195} IC according to eq.(6.2). This is clearly not feasible and the alternative approaches analysed below could help the researcher to reduce the computational burden.

Simulated Annealing (SA)

This algorithm provides a local search for the minimum (or maximum) of a function, in our case eq.(6.2). The concept is originally based on the manner in which liquids freeze or metals recrystalize in the process of annealing. In an annealing process a melt, initially at high temperature and disordered, is slowly cooled so that the system at any time is approximately in thermodynamic equilibrium. As cooling proceeds, the system becomes more ordered and approaches a 'frozen' ground state. The analogy to an optimisation problem is as follows: the current state of the thermodynamic system is analogous to the current solution to the optimisation problem, the energy equation for the thermodynamic system is analogous to the objective function, and the ground state is analogous to the global optimum. An early application of simulated annealing in econometrics is the work of Goffe et al.   (1994) who suggested that simulated annealing could be used to optimise the objective function of various econometric estimators.

Below, we give a description of the algorithm together with the necessary arguments that illustrate its validity in our context. We describe the operation of the algorithm when the domain of the function (information criterion) is the set of binary strings, i.e., $\{\mathcal{I} = (\mathcal{I}_1, \ldots, \mathcal{I}_N)' | \mathcal{I}_i \in \{0, 1\}\}$.

Each step of the algorithm works as follows starting from an initial string $\mathcal{I}_0$.

1. Using $\mathcal{I}_i$ choose a neighboring string at random, denoted $\mathcal{I}_{i+1}^*$. We discuss the definition of a neighborhood below.

2. If $IC(\mathcal{I}_i) > IC(\mathcal{I}_{i+1}^*)$, set $\mathcal{I}_{i+1} = \mathcal{I}_{i+1}^*$. Else, set $\mathcal{I}_{i+1} = \mathcal{I}_{i+1}^*$ with probability $e^{(IC(\mathcal{I}_i^*) - IC(\mathcal{I}_{i+1}))/T_i}$ or set $\mathcal{I}_{i+1} = \mathcal{I}_i$ with probability $1 - e^{(IC(\mathcal{I}_i^*) - IC(\mathcal{I}_{i+1}))/T_i}$.

Heuristically, the term T_i gets smaller making it more difficult, as the algorithm proceeds, to choose a point that does not decrease $IC(.)$. The issue of the neighborhood is extremely relevant. What is the neighborhood?

Intuitively, the neighborhood could be the set of strings that differ from the current string by one element of the string. But this may be too restrictive. We can allow the algorithm to choose at random, up to some maximum integer (say h), the number of string elements at which the string at steps i and $i+1$ will differ. So the neighborhood is all strings with up to h different bits from the current string. Another issue is when to stop the algorithm. There are a number of alternatives in the literature. We have chosen to stop the algorithm if it has not visited a string with lower $IC(.)$ than the current minimum for a prespecified number of steps (B_v) (Steps which stay at the same string do not count) or if the number of overall steps exceeds some other prespecified number (B_s). All strings visited by the algorithm are stored and the best is chosen at the end rather than the final one.

The simulated annealing algorithm has been proven by Hajek (1998) to converge asymptotically, i.e. as $i \to \infty$, to the maximum of the function as long as $T_i = T_0/ln(i+1)$ for some sufficiently large T_0. In particular, for almost sure convergence to the minimum it is required that $T_0 > d^*$, where d^* denotes the maximum depth of all local minima of the function $IC(.)$. Heuristically, the depth of a local minimum, $\mathcal{I}_1$, is defined as the smallest number $E > 0$ such that the function exceeds $IC(\mathcal{I}_1) + E$ during its trajectory from this minimum to any other local minimum, $\mathcal{I}_2$, for which $IC(\mathcal{I}_1) > IC(\mathcal{I}_2)$.[1]

This condition needs to be made specific for the problem at hand. We thus need to discuss possible strategies for determining d^* for model searches using information criteria. It is reasonable to assume that the space of models searched via information criteria only includes models with a prespecified maximum number of variables, otherwise problems caused by the lack of degrees of freedom will arise. Then, a possible upper limit for d^* is $2L(\mathcal{I}_B) - 2L(\mathcal{I}_A)$ where $L(\mathcal{I}_A)$ is the likelihood associated with a regression containing just a constant term and $L(\mathcal{I}_B)$ is the likelihood associated with a regression containing the maximum allowable number of variables. Of course, there are many possible sets of variables that contain the maximum allowable number of variables. For this reason we remove the penalty terms and focus on likelihoods. This makes it more likely that $-2L(\mathcal{I}_B)$, for some random $\mathcal{I}_B$ that specifies use of the maximum allowable number of variables, is a lower bound for the optimum value taken by the information criterion.

Genetic Algorithm (GA)

The motivating idea of genetic algorithms is to start with a population of binary strings which then evolve and recombine to produce new populations with 'better' characteristics, i.e. lower values for the information criterion. We start with an initial population represented by a $N \times m$ matrix made up of 0's and 1's. Columns represent strings. m is the chosen size of the population. The theory of genetic algorithms suggests that the composition of the initial population does not matter. Hence, this is generated randomly.

Denote this population matrix by P_0. The genetic algorithm involves defining a transition from $\mathbf{P}_i$ to $\mathbf{P}_{i+1}$. Following Kapetanios (2007), the algorithm could be described in the following steps:

1. For $\mathbf{P}_i$ create a $m \times 1$ 'fitness' vector, $\mathbf{p}_i$, by calculating for each column of $\mathbf{P}_i$ its 'fitness'. The choice of the 'fitness' function is completely open and depends on the problem. For our purposes it is the opposite of the information criterion. Normalise $\mathbf{p}_i$, such that its elements lie in $(0,1)$ and add up to 1. Denote this vector by $\mathbf{p}_i^*$. Treat $\mathbf{p}_i^*$ as a vector of probabilities and re-sample m times out of $\mathbf{P}_i$ with replacement, using the vector $\mathbf{p}_i^*$ as the probabilities with which each string will be sampled. So 'fit' strings are more likely to be chosen. Denote the re-sampled population matrix by $\mathbf{P}_{i+1}^1$.

2. Perform cross over on $\mathbf{P}_{i+1}^1$. For cross over we do the following: Arrange all strings in $\mathbf{P}_{i+1}^1$, in pairs (assume that m is even) where the pairings are randomly drawn. Denote a generic pair by $(a_1^\alpha, a_2^\alpha, \ldots, a_N^\alpha), (a_1^\beta, a_2^\beta, \ldots, a_N^\beta)$. Choose a random integer between 2 and $N-1$. Denote this by j. Replace the pair by the following pair: $(a_1^\alpha, a_2^\alpha, \ldots, a_j^\alpha, a_{j+1}^\beta, \ldots, a_N^\beta), (a_1^\beta, a_2^\beta, \ldots, a_j^\beta, a_{j+1}^\alpha, \ldots, a_N^\alpha)$.

---

[1] A trajectory from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a set of strings, $\mathcal{I}_{11}, \mathcal{I}_{12}, \ldots, \mathcal{I}_{1p}$, such that (i) $\mathcal{I}_{11} \in N(\mathcal{I}_1)$, (ii) $\mathcal{I}_{1p} \in N(\mathcal{I}_2)$ and (iii) $\mathcal{I}_{1i+1} \in N(\mathcal{I}_{1i})$ for all $i = 1, \ldots, p$, where $N(\mathcal{I})$ denotes the set of strings that make up the neighborhood of $\mathcal{I}$.

Perform cross over on each pair with probability p_c. Denote the new population by $\mathbf{P}^2_{i+1}$. Usually p_c is set to some number around 0.5-0.6.

3. Perform mutation on $\mathbf{P}^2_{i+1}$. This amounts to flipping the bits (0 or 1) of $\mathbf{P}^2_{i+1}$ with probability p_m. p_m is usually set to a small number, say 0.01. After mutation the resulting population is $\mathbf{P}_{i+1}$.

These steps are repeated a pre-specified number of times (B_g). Each set of steps is referred to as generation in the genetic literature. If a string is to be chosen this is the one with maximum fitness. For every generation we store the identity of the string with maximum 'fitness'. Further this string is allowed to remain intact for that generation. So it gets chosen with probability one in step 1 of the algorithm and does not undergo neither cross-over nor mutation. At the end of the algorithm the string with the lowest information criterion value over all members of the populations and all generations is chosen.

One can think of the transition from one string of maximum fitness to another as a Markov Chain. So this is a Markov Chain algorithm. In fact, the Markov chain defined over all possible strings is time invariant but not irreducible as at least the $m - 1$ least fit strings will never be picked. To see this note that in any population there will be a string with more fitness than that of the $m - 1$ worst strings. There has been considerable work on the theoretical properties of genetic algorithms. Hartl and Belew (1990) have shown that with probability approaching one, the population at the n-th generation will contain the global maximum as $n \to \infty$. Perhaps the most relevant result from that work is Theorem 4.1 of Hartl and Belew (1990). This theorem states that as long as (i) the sequence of the maximum fitnesses in the population across generations is monotonically increasing, and (ii) any point in the model space is reachable from any other point by means of mutation and cross-over in a finite number of steps then the global maximum will be attained as $n \to \infty$.

Both these conditions hold for the algorithm described above. The first condition holds by the requirement that the string with the maximum fitness is always kept intact in the population. The second condition holds since any string of finite length can be obtained from another by cross-over and mutation with non-zero probability in a finite number of steps. For more details on the theory of genetic algorithms see also Morinaka et al. (2001).

MC^3

This algorithm is similar to simulated annealing for the construction of its steps. This similarity is, in fact, the main reason why we consider Bayesian methods here. The MC^3 algorithm defines a search path in the model space just like the simulated annealing algorithm we considered above. As a result we refer to the setup of the previous section to minimise duplication for the exposition. The difference between SA and MC^3 is the criterion used to move from one string to the other at step i. Here, the Bayes factor for string (model) $i + 1$ versus string (model) i is used. This is denoted by $B_{i+1,i}$. The chain moves to the $i + 1$ string with probability $min(1, B_{i+1,i})$. This is again a Metropolis-Hastings type algorithm. The Bayes factor we use following Fernandez et al. (2001) is given by,

$$
B_{i+1,i} = \left(\frac{g_{0i+1}}{g_{0i+1} + 1} \right)^{k_{i+1}/2} \left(\frac{g_{0i} + 1}{g_{0i}} \right)^{k_i/2}
$$
$$
\left(\frac{\frac{1}{g_{0i+1}} RSS_i + \frac{g_{0i}}{g_{0i}+1} TSS}{\frac{1}{g_{0i+1}+1} RSS_{i+1} + \frac{g_{0i+1}}{g_{0i+1}+1} TSS} \right)^{(T-1)/2},
$$

(6.3)

where RSS_i is the sum of squared residuals of the i-th model, TSS is the sum of the squared deviations from the mean for the dependent variable, k_i is the number of variables in model i and g_{0i} is a model specific constant relating to the prior relative precision. The results of Fernandez et al. (2001) suggest that for consistent model selection g_{0i} should be set to $1/T$. This is associated with prior 'a' in the terminology of subsection 4.2 of Fernandez et al. (2001). More details may be found in Fernandez et al. (2001). Our

chosen model is the model that minimises the information criterion among all models visited by the MC^3 algorithm. Given the results of Annex A.3 of Fernandez et al. (2001) concerning the asymptotic equivalence between consistent information criteria and the Bayes factor in (6.3) we find our approach justified.

Sequential Testing (ST)

A general regression specification is considered and tested for misspecification using a battery of specification tests such as tests for residual autocorrelation and ARCH and tests for structural breaks. Then, a sequential testing procedure is used to remove insignificant regressors from this specification making sure that the resulting specifications are acceptable using misspecification tests. This algorithm provides a tractable formalisation of the general-to-specific methodology advocated by David Hendry and his co-authors and discussed in some detail in a number of papers such as, e.g., Hendry (1995) and Hendry (1997) (see also Bruggemann et al. (2003) for an application of this methodology to model reduction in VAR processes).

A detailed description of the algorithm we use is given in steps A-H of Hoover and Perez (1999). The only modifications to this algorithm are as follows: (i) All possible search paths, rather than only 10, are considered. (ii) In step B(d) we use $CUSUM^2$ instead of Chow as a stability test. (iii) No out-of-sample evaluation is undertaken, since this would change the information set for the other algorithms. We try two versions of this algorithm for two different significance levels for all the tests involved (5% and 1%) using AIC.

6.2.3 Variable Reduction Methods

Factor models are very useful in the construction of (Cyclical) Composite Indicators, see Chapters 9 and 13 for detailed information. In this Chapter and in Chapter 12 we assess their forecasting performance related to specific benchmarks.

Principal Components (PC)

The most widely used class of data-rich forecasting methods are factor methods. Factor methods have been at the forefront of developments in forecasting with large data sets and in fact started this literature with the influential work of Stock and Watson (2002a). The defining characteristic of most factor methods is that relatively few summaries of the large data sets are used in the forecasting equation, which thereby becomes a standard forecasting equation as it only involve a few variables. The assumption is that the co-movements across the indicator variables x_t, where $x_t = (x_{1,t} \cdots x_{N,t})'$ is a vector of dimension $N \times 1$, can be captured by a $r \times 1$ vector of unobserved factors $F_t = (F_{1,t} \cdots F_{r,t})'$, i.e.,

$$\tilde{x}_t = \Lambda' F_t + e_t, \tag{6.4}$$

where $\tilde{x}_t$ may be equal to x_t or may involve other variables such as, e.g., lags and leads of x_t and Λ is a $r \times N$ matrix of parameters describing how the individual indicator variables relate to each of the r factors, which we denote with the terms 'loadings'. In (6.4) e_t represents a zero-mean $I(0)$ vector of errors that represent for each indicator variable the fraction of dynamics unexplained by F_t, the 'idiosyncratic components'. The number of factors is assumed to be small, meaning $r < \min(N, T)$. So, implicitly, in (6.1) $\alpha' = \tilde{\alpha}' \Lambda \tilde{x}_t$, where $F_t = \Lambda \tilde{x}_t$, which means that a small, r, number of linear combinations of $\tilde{x}_t$ represent the factors and act as the predictors for y_t which denotes the target variable. The main difference between different factor methods relates to how Λ is estimated.

The use of principal components (PC) for the estimation of factor models is, by far, the most popular factor extraction method. It has been popularised by Stock and Watson (2002a) and Stock and Watson (2002b), in the context of large data sets, although the idea had been well established in the traditional multivariate

statistical literature. The method of principal components is simple. Estimates of Λ and the factors F_t are obtained by solving,

$$V(r) = \min_{\Lambda, F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{x}_{i,t} - \lambda_i' F_t)^2, \tag{6.5}$$

where λ_i is a $r \times 1$ vector of loadings that represent the N columns of $\Lambda = (\lambda_1 \cdots \lambda_N)$. One, non-unique, solution of (6.5) can be found by taking the eigenvectors corresponding to the r largest eigenvalues of the second moment matrix $X'X$, which then are assumed to represent the rows in Λ, and the resulting estimate of Λ provides the forecaster with an estimate of the r factors $\hat{F}_t = \hat{\Lambda}\tilde{x}_t$. To identify the factors up to a rotation, the data are usually normalized to have zero mean and unit variance prior to the application of principal components; see Stock and Watson (2002a) and Bai (2003).

PC estimation of the factor structure is essentially a static exercise as no lags or leads of x_t are considered. One alternative is dynamic principal components, which, as a method of factor extraction, has been suggested in a series of papers by Forni, Hallin, Lippi and Reichlin (see, e.g., Forni et al. (2000) among others). Dynamic principal components are extracted in similar fashion to static principal components but , instead of the second moment matrix, the spectral density matrices of the data at various frequencies are used. These are then used to construct estimates of the common component of the data set which is a function of the unobserved factors. This method uses leads of the data and as a result its application to forecasting has been slow for obvious reasons. Recent work by the developers of the method has addressed this issue (see, e.g., Forni et al. (2005)). However, evidence suggests that static PC are a more effective and robust technique for forecasting and we will therefore, focus on this in our evaluation.

Partial Least Squares (PLS)

Partial least squares (PLS) is a relatively new method for estimating regression equations, introduced in order to facilitate the estimation of multiple regressions when there is a large, but finite, amount of regressors.[2] The basic idea is similar to principal component analysis in that factors or components, which are linear combinations of the original regression variables, are used, instead of the original variables, as regressors. A major difference between PC and PLS is that, whereas in PC regressions the factors are constructed taking into account only the values of the x_t variables, in PLS, the relationship between y_t and x_t is considered as well in constructing the factors. PLS regression does not seem to have been explicitly considered for data sets with a very large number of series, i.e., when N is assumed in the limit to converge to infinity.

There are a variety of definitions for PLS and accompanying specific PLS algorithms that inevitably have much in common. A conceptually powerful way of defining PLS is to note that the PLS factors are those linear combinations of x_t, denoted by Υx_t, that give maximum covariance between y_t and Υx_t while being orthogonal to each other. Of course, in analogy to PC factors, an identification assumption is needed, to construct PLS factors, in the usual form of a normalization.

A simple algorithm to construct k PLS factors is discussed among others, in detail, in Helland (1990). Assuming for simplicity that y_t has been demeaned and x_t have been normalized to have zero mean and unit variance, a simplified version of the algorithm is given below

1. Set $u_t = y_t$ and $v_{i,t} = x_{i,t}$, $i = 1, ...N$. Set $j = 1$.

2. Determine the $N \times 1$ vector of indicator variable weights or loadings $w_j = (w_{1j} \cdots w_{Nj})'$ by computing individual covariances: $w_{ij} = Cov(u_t, v_{it})$, $i = 1, ..., N$. Construct the j-th PLS factor by taking the linear combination given by $w_j' v_t$ and denote this factor by $f_{j,t}$.

---

[2] Herman Wold and co-workers introduced PLS regression between 1975 and 1982, see, e.g., Wold (1982). Since then it has received much attention in a variety of disciplines, especially in chemometrics, outside of economics.

3. Regress u_t and $v_{i,t}$, $i = 1,...,N$ on $f_{j,t}$. Denote the residuals of these regressions by $\tilde{u}_t$ and $\tilde{v}_{i,t}$ respectively.

4. If $j = k$ stop, else set $u_t = \tilde{u}_t$, $v_{i,t} = \tilde{v}_{i,t}$ $i = 1,..,N$ and $j = j + 1$ and go to step 2.

This algorithm makes clear that PLS is computationally tractable for very large data sets. Once PLS factors are constructed y_t can be modeled or forecasted by regressing y_t on $f_{j,t}$ $j = 1,...,k$. Helland (1988) and Helland (1990) provide a general description of the partial least squares (PLS) regression problem. Helland (1988) shows that the estimates of the coefficients α in a regression of y_t on x_t, as in (6.1), obtained implicitly via PLS Algorithm and a regression of y_t on $f_{j,t}$ $j = 1,...,k$, are mathematically equivalent to,

$$\hat{\beta}_{PLS} = V_k(V_k'X'XV_k)^{-1}V_k'X'y, \tag{6.6}$$

with $V_{k_1} = (X'y \quad X'XX'y \quad \cdots \quad (X'X)^{k-1}X'y)$, $X = (x_1\cdots x_T)'$ and $y = (y_1\cdots y_T)'$. Thus, (6.6) suggests that the PLS factors that result from the PLS Algorithm span the Krylov subspace generated by $X'X$ and $X'y$, resulting in valid approximations of the covariance between y_t and x_t.

Bayesian Shrinkage Regression (BR)

Bayesian regression is a standard tool for providing inference for α in (6.1) and there exists a large variety of approaches for implementing Bayesian regression. We will provide a brief exposition of this method. A starting point is the specification of a prior distribution for α. Once this is in place standard Bayesian analysis proceeds by incorporating the likelihood from the observed data to obtain a posterior distribution for α which can then be used for a variety of inferential purposes, including, of course, forecasting.

A popular and simple implementation of Bayesian regression results in a shrinkage estimator for α in (6.1) given by,

$$\hat{\beta}_{BRR} = (X'X + vI)^{-1}X'y, \tag{6.7}$$

where $X = (x_1,...,x_T)'$, $y = (y_1,..,y_T)'$ and v is a shrinkage scalar parameter. The shrinkage estimator (6.7) shrinks the OLS estimator, given by $(X'X)^{-1}X'y$ towards zero, thus enabling a reduction in the variance of the resulting estimator. This is a major feature of Bayesian regression that makes it useful in forecasting when large data sets are available. This particular implementation of Bayesian regression implies that elements of α are small but different from zero ensuring that all variables in x_t are used for forecasting. In this sense, Bayesian regression can be linked to other data-rich approaches. When a certain factor structure is assumed in the data, Bayesian regression through (6.7) will forecast y_t by projecting it on a weighted sum of all N principal components of X, with decaying weights, instead of projecting it on a limited number of r principal components with equal weights as in PC regression; see Del Mol et al. (2006).

6.2.4 From Large to Small and Medium Datasets

Average Correlation

A simple approach to choose some of the variables from a large dataset is to categorise them based on their nature and select the most representative variable within each group. Table 6.B.1 shows starting set of monthly variables, divided in 22 groups. At any round in the cross-validation algorithm (that follows later) we estimate the cross-correlations in each group and we select the variable that presents the highest average correlation with the other variables of the same group. Since this operation is repeated at the beginning of each period (month) we refer to the method as Average Dynamic ($AVGDYN$).

In a recent study Bailey, Kapetanios and Pesaran (2012) suggest ways to estimate the cross-sectional expo-
nent. Consider the factor model as in eq. (6.4). The main idea is that we allow the loadings r to be fixed in N,
but we assume that only N^α of the N factor loadings are individually important. The focus of their analysis is
on α, the cross-sectional exponent. Specifically they consider,

$$r_{ik} = v_{ik} \text{ for } i = 1, 2, ..., \lfloor N^{\alpha_k} \rfloor ,$$
$$r_{ik} = c_k \rho_k^{i - \lfloor N^{\alpha_k} \rfloor + 1} \text{ for } i = \lfloor N^{\alpha_k} \rfloor + 1, \lfloor N^{\alpha_k} \rfloor + 2, ..., N, \tag{6.8}$$

for $k =$1, 2, ...,m, where $\lfloor \cdot \rfloor$ denotes the integer part, $1/2 < \alpha_k <$1, $|\rho_k| < 1$, c_k is a finite constant and
$v_{ik} \sim iid\left(\mu_{v_k}, \sigma_{v_k}^2\right)$ with $\mu_{v_k} \neq 0$ and $\sigma_{v_k}^2 > 0$. The authors provide the theoretical background and
inference to estimate $\alpha = \max_k \alpha_k$.

A simple consistent estimator is given by,

$$\widehat{\alpha_1} = 1 + \frac{\ln\left(\widehat{\sigma}_{\overline{x}}^2\right)}{2\ln(N)}, \tag{6.9}$$

where $\widehat{\sigma}_{\overline{x}}^2 = T^{-1} \sum_{t=1}^{T} \left(\overline{x}_t - \overline{x}\right)^2$ and $\overline{x} = T^{-1} \sum_{t=1}^{T} \overline{x}_t$. Two bias adjusted versions of the above follow.

$$\widehat{\alpha_2} = \widehat{\alpha_1} - \frac{\widehat{\overline{\sigma}_N^2}}{2\ln(N) N \widehat{\sigma}_{\overline{x}}^2}, \tag{6.10}$$

$$\widehat{\alpha_3} = \widehat{\alpha_1} - \frac{\widehat{\overline{\sigma}_N^2}}{2\ln(N) N \widehat{\sigma}_{\overline{x}}^2} \left(1 + \frac{\widehat{\overline{\sigma}_N^2}}{N \widehat{\sigma}_{\overline{x}}^2}\right), \tag{6.11}$$

where $\widehat{\overline{\sigma}_N^2} = N^{-1} \sum_{i=1}^{N} \widehat{\sigma}_i^2$, $\widehat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \widehat{u}_{it}^2$, $\widehat{u}_{it} = x_{it} - \widehat{\delta}_i \widetilde{x}_t$, $\widetilde{x}_t = \widehat{\sigma}_{\overline{x}}^{-1}\left(\overline{x}_t - \overline{x}\right)$ and $\widehat{\delta}_i$ denotes the OLS
estimator of the regression coefficient of x_{it} on $\widetilde{x}_t$. The last estimator considered is the ρ-adjusted estimator
and is defined as,

$$\widehat{\alpha_4} = \widehat{\alpha_3} + \frac{\ln\left(\kappa^2\right)}{2\ln(N)}, \tag{6.12}$$

where $\kappa^2 = \sum_{i=1}^{m} \mu_{iv}^2 \sigma_{ir}^2$. The authors suggest ways to estimate the above; for further details see Bailey,
Kapetanios and Pesaran (2012). Once an estimate for α is obtained then we rank the variables according to
the absolute value of their loadings (max to min) and we select the first $\lfloor N^{\alpha_j} \rfloor$ variables for $j = \{1, 2, 3, 4\}$.
We do the above procedure at the beginning of each round in the cross-validation forecasting algorithm and
in that sense the set of predictors changes dynamically.

6.2.5 Parameters Setup and Normalisation

For the simulated annealing and genetic algorithms we use the same (default) values as in Kapetanios (2007),
i.e. in the simulated annealing $h = 1$, $B_v = 500$, $B_s = 5000$, $T_0 = 10$, in the genetic algorithm $m = 200$,
$B_g = 200$, $p_c = 0.6$ and $p_m = 0.1$. We allow the max counter of convergence iterations to be 10 and 500
times. In PC and PLS we examine cases with 1 and 3 factors and in BR we use $v = 0.5N$ and $v = 2N$ as
shrinkages.

In all heuristics we have used the data as is. However, in PC and PLS we have normalised the regressors to zero mean and unit variance series.

6.3 Forecasting Exercise and Data Description

Let us assume that we have obtained x_t^0 using one of the methods previously described. We perform a forecasting exercise using the projection method as described in Stock and Watson (2002a). This method, also known as direct approach, is more robust in the presence of possible model mis-specification. The forecasts are given by,

$$\widehat{y}_{t+h}^f = \widehat{\beta}^{h\prime} x_t^0, \tag{6.1}$$

where $\widehat{\beta}^h$ is obtained by regressing y_t on x_{t-h}^0 and h denotes the forecast horizon. At first, we set the $steps$ ahead in the forecast, h. Then, we specify the evaluation period, $Eval$, and we omit h observations completely out of the sample. This allows us to end up with a number of $Eval$ forecasts for any given step h. A summary of the pseudo out-of-sample forecasting algorithm follows.

1. Use an initial sample of T_1 observations ($T_1 = T - Eval - h$). The number of the *initial* set of predictors could be 195 or could vary depending on the dataset method used. For AVGDYN and $\widehat{\alpha_j}$, $j = \{1, 2, 3, 4\}$, methods the new dataset of predictors changes dynamically.

2. With any method described in this section obtain $x_t^{0\prime}$, $t = 1, 2, ..., T_1$,

3. For $j = 1, 2, ..., h$ steps regress y_t on $x_{t-h}^{0\prime}$ and obtain $\widehat{\beta}^h = \left(\widehat{\beta}^1, ..., \widehat{\beta}^h \right)'$,

4. Calculate the forecasts of $\widehat{y}_{t+h}^f$ using $x_t^{0\prime}$ and $\widehat{\beta}^h$, hence $\widehat{y}^f = \left(\widehat{y}_1^f, ..., \widehat{y}_h^f \right)'$,

5. Repeat the whole procedure increasing the initial sample T_1 to $T_l = T_{l-1} + 1$ until $T_l = T - h$.

At the end of this process we will have gathered a number of $Eval$ forecast values for any step h. The forecast error is then calculated as,

$$\widehat{e}_{t+h}^f = y_{t+h} - \widehat{y}_{t+h}^f, \tag{6.2}$$

and the statistics of interest can be computed. We are particularly interested in the Root Mean Squared Forecast Error[3] defined as:

$$RMSFE_h = \sqrt{\frac{1}{Eval} \sum_{j=1}^{Eval} \left(e_{t+h,j}^f \right)^2}. \tag{6.3}$$

The regressors (predictors) data we have available consists of 195 monthly variables (source: Eurostat, PEEIs, the Eurostat labels can be found in the annex) spanning from Jan. 1996 to Mar. 2009. The dataset is the same used in Foroni and Marcellino (2011) and it contains a large universe of variables that are potentially useful instruments in forecasting key macroeconomic variables in the Euro Area such as those described below. Furthermore, in the spirit of Stock and Watson (2002a) we have transformed the series for stationarity using first differences or $\log$ differences appropriately (although notice in Table 6.C.1, some of the variables remained unchanged). Hence, the resulting data used in the forecast exercise contains growth rates from Feb. 1996 to Mar. 2009 (inclusive).

In particular, our list of dependent variables consists of growth rates for the:

---

[3]Diebold-Mariano test results are available on request from the authors and are not reported here due to length limitations.

- Quarterly GDP (EA-16), seasonally adjusted (source: ECB/Eurostat),

- Quarterly Consumption (EA-17), seasonally adjusted (source: ECB/Eurostat),

- Monthly Industrial Production Growth Rate to previous period, seasonally adjusted (source: ECB/Eurostat), and the

- Monthly HICP (EA-16) (source: Eurostat).

In the cases of consumption, GDP and HICP we calculate the growth rate using the $\log$ transformation, hence,

$$g_t\left(y_t\right) = \ln\left(\frac{y_t}{y_{t-1}}\right), \tag{6.4}$$

whereas in the cases of the industrial production we use the rates as given.

The seasonal adjustments are made by the ECB/Eurostat and the series are provided transformed. The dates are as in the regressors case, Jan. 1996 to Mar. 2009, but due to the necessary transformations for stationarity we end up using Feb. 1996 to Mar. 2009. Tables 6.B.1 and 6.C.1 provide the labels of all regressors used along with the relevant transformations.

An obvious problem that arises using quarterly regressands and monthly regressors is how to cope with the frequency mismatch. To avoid this problem we are assess two different approaches to translate the monthly regressors to quarters:

1. Using the quarters averages, i.e. averaging out the three months in the same quarter,

2. Using the third month in each quarter only.

We refer to above methods as $3MA$ and $3MObs$ respectively. As soon as we have made the above translation we impose the stationarity transformations as before.

Regarding the quarterly data we always start the forecasting algorithm using the first 32 observations, i.e. Q4. 1996 to Q3. 2004. This is a initial sample for the optimisation algorithms to converge to a solution. In the $4 - steps$ ahead forecast we perform the forecast evaluation 16 times (hence for the 16 subsequent periods). Using the monthly data we start with the first 60 observations, i.e. Feb. 1996 to Jan. 2001, with $12 - steps$ ahead forecast[4] and an evaluation period of 84. In all cases we use recursive looping and we leave 4 and 12 observations out of sample as in step 1 of the forecasting algorithm described above.[5]

6.3.1 Benchmark: AR(1)

In order to compare the predictive ability of the methods suggested here, we use as a benchmark an autoregressive model of order 1 described as,

$$y_t = \phi y_{t-1} + \varepsilon_t,\ t = 1, 2, ..., T. \tag{6.5}$$

The forecasts in this case are computed recursively.[6] The forecast errors and RMSFE are computed as before (eq.(6.2) and eq.(6.3)).

---

[4]However, for reasons of brevity only $h = \{1, 3, 6, 12\}$ are reported.

[5]Note that one could also produce monthly updates of the quarterly forecasts. This exercise is considered in Bulligan, Marcellino and Venditti (2011), whose results are qualitatively similar to ours, confirming the relevance of variable selection.

[6]Results using the direct approach are also available on request. Comparing the recursive AR forecasting and the direct method, the first led to better forecast results on average and this is why it is reported here.

6.4 Discussion of Results

For reasons of notational convenience the method minimising a specific criterion will be denoted by $Method_{IC}$, e.g. the simulated annealing algorithm that uses AIC as objective function is denoted SA_{AIC} and the sequential testing at 5% level is denoted ST_{AIC}^5. For the PC and PLS the number of factors is displayed inside the brackets, while for BR inside the brackets there is the shrinkage parameter.[7]

6.4.1 Forecasting the GDP Growth Rate

Tables 6.A.1 to 6.A.3 present the forecasting exercise when we translate monthly regressors to quarterly data using the 3-months averaging (i.e. 3MA) methodology. In Table 6.A.1 we compare the forecasting performance of all methods using the full dataset and the small dataset of 22 variables given by AVGDYN method. At first, all variable reduction methods fail to provide forecasts with smaller errors compared to the AR(1) benchmark. However, we see that the variable selection methods outperform the AR(1) in 2 to 4 steps ahead. For one-step ahead, only MC_{AIC}^3 and GA_{BIC} have a relative RMSFE smaller than unity. In particular, the actual RMSFE of the AR(1) is 0.004 and for the above methods the relative RSMFE is 0.971 and 0.938 respectively.

Using the AVGDYN dataset, the performance of all methods improves for horizons 2 to 4. However, one-step ahead only GA_{HQ} and MC_{HQ}^3 return better forecasts with relative RMSFE of 0.955 and 0.999 respectively.

In Tables 6.A.2 and 6.A.3 where datasets of predictors are derived using the estimates of the cross-sectional exponent $\lfloor N^{\widehat{\alpha}_j} \rfloor$, $j = \{1, 2, 3, 4\}$, we see that the forecasting performance of the variable selection methods has improved. Using $\widehat{\alpha}_1$ we have that for $h = 1$ MC_{AIC}^3, ST_{AIC}^1, SA_{BIC} and MC_{BIC}^3 provide a relative RMSFE of 0.755, 0.763, 0.816 and 0.799 respectively. The most robust method is ST_{AIC}^1 which has a relative RSMFE of 0.576, 0.57, 0.551 for $h =$ 2, 3 and 4 respectively. The same methods still outperform the benchmark using $\widehat{\alpha}_2$. However, MC_{BIC}^3 is now the most robust with relative RSMFE of 0.984, 0.505, 0.680, 0.686 for $h =$ 1, 2, 3 and 4 steps ahead. Finally, using $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ we see again that MC_{AIC}^3, SA_{BIC}, GA_{BIC}, MC_{BIC}^3 and MC_{HQ}^3 are the best variable selection methods as they have a relative RMSFE less than unity in most of the cases. On the contrary, all variable reduction methods fail to compete with the benchmark using all $\lfloor N^{\widehat{\alpha}_j} \rfloor$, $j = \{1, 2, 3, 4\}$ datasets.

Tables 6.A.4 to 6.A.6 present the results when translating the monthly to quarterly variables using the third month in each quarter (i.e. 3MObs). As before, none of the variable reduction methods is better compared to the AR(1) benchmark. Using the original dataset of predictors only GA_{BIC} provides better forecasts with relative RMSFE of 0.837, 0.88, 0.613 and 0.723 for 1, 2, 3 and 4 steps ahead respectively. Other methods that include ST_{AIC}^1, ST_{AIC}^5, SA_{BIC}, MC_{BIC}^3, GA_{HQ} and MC_{HQ}^3 return forecasts with smaller error compared to the AR(1) only in 2 to 4 steps ahead. Using the AVGDYN method for selecting predictors all variable selection methods outperform the benchmark for $h = 2, 3, 4$. Using the cross-sectional exponent estimates methodology, we have that the forecasting ability of the variable selection methods is improved and the results further enhance the robustness of the above mentioned methods.

6.4.2 Forecasting the Consumption Growth Rate

In Tables 6.A.7 to 6.A.9 we present the results for the final consumption expenditure growth rate using 3MA transformation. The cases of the large and the small AVGDYN datasets in Table 6.A.7 clearly show how the small dataset helps the forecasting performance. All variable selection methods outperform the benchmark with extremely small relative RMSFE. The best method for step 1 is GA_{HQ} with a relative RSMFE of 0.372 and for all subsequent steps we have ST_{AIC}^1 with relative RSMFE equal to 0.643, 0.616 and 0.659. In the variable reduction methods there is not significant difference. PC and PLS methods perform better than the

---

[7]All Diebold-Mariano results are available on request. In monthly data, results for steps $h = 1, ..., 12$ are also available.

benchmark in 1 step ahead but they cannot compete with the previously mentioned models. In Tables 6.A.8 to 6.A.9 we generally see that MC^3_{AIC}, ST^1_{AIC}, GA_{BIC} and MC^3_{BIC} are robust across all steps ahead. The variable reduction methods are slightly improved but still underperform compared to variable selection methods.

Using the 3MObs transformation to translate the predictors from monthly to quarterly, see Tables 6.A.10 to 6.A.12, we reach a similar qualitative result as before. In the full dataset, ST^1_{AIC} and GA_{BIC} outperform the benchmark in all steps ahead whereas when using the small dataset suggested by the AVGDYN method, all variable selection models provide forecasts with smaller error. Using the datasets suggested by the estimates of the cross-sectional exponent we see that MC^3_{AIC}, SA_{BIC}, GA_{BIC}, MC^3_{BIC} and MC^3_{HQ} perform better than the benchmark and the variable reduction methods.

6.4.3 Forecasting the Industrial Production Growth Rate

Tables 6.A.13 to 6.A.15 present the forecasting results for the monthly industrial production growth. Comparing the results of the forecasting exercise using the original large dataset of predictors with the AVGDYN and the cross-sectional exponent estimates datasets, we see that there is improvement in the forecasting power. Specifically, using the AVGDYN method almost all variable selection models provide better forecasts compared to the benchmark in 1 and 3 steps ahead. The variable reduction models are also improved using this dataset. A representative example is the ST^1_{AIC} where its relative RMSFE is 0.972 from 1.037 in step 1.

Using the estimates of the exponent of the cross-sectional averages we see that GA_{BIC} and MC^3_{BIC} are assisted in the variables selection procedure as their forecasts are now better. The variable reduction methods are slightly improved as before.

6.4.4 Forecasting the HICP Growth Rate

The last dependent variable we are concerned with in this study is the growth of the Harmonised Price Index. In Tables 6.A.16 to 6.A.18 we see that using the full original dataset the ST^1_{AIC}, SA_{BIC}, GA_{BIC}, MC^3_{BIC}, GA_{HQ} and MC^3_{HQ} models outperform the benchmark for 1, 3, 6 and 12 steps ahead. The best model for 1 step ahead is the ST^1_{AIC} with a relative RMSFE of 0.815. For 3 and 12 steps ahead MC^3_{HQ} performs better on average with relative RMSFE of 0.908 and 0.797 respectively. PC(3) and PLS(3) perform better among the variable reduction methods but cannot compete with the above mentioned variable selection models.

Using the small AVGDYN dataset we clearly see how the forecasting performance of the variable selection methods has improved. All methods outperform the benchmark while the forecasts made by the various variable reduction models are slightly worse than before. Using the datasets indicated by the estimates of the cross-sectional exponent we see that PC(3), PLS(1) and PLS(3) now provide more accurate forecasts compared to the benchmark, however the variable selection methods still provide results with smaller error on average. In particular, MC^3_{AIC} and all other optimisation methods that minimise BIC and HQ outperform the benchmark.

In Tables 6.A.19 to 6.A.21 we present the results adding one more variable in the original large set of predictors which is the first order autoregressive series of HICP growth. We do so as the majority of applied researches often include past lag(s) of the dependent variable. However, we find that our qualitative conclusions hold.

6.5 Conclusions

In this chapter we evaluate how the forecasting performance of various model selection and model reduction techniques is affected by the use of small, medium and large sets of predictors. We initially start our

experiments using a very large set of 195 variables and then we create five additional subsets.

Firstly, a small set of predictors is formed by dividing all variables in categories and selecting the most representative variable within each group.

Then, since recent research by Bailey, Kapetanios and Pesaran (2012) suggests that N^α out of all predictors are individually important (α here is called the cross-sectional exponent), we employ four estimates for α introduced in the above paper and we end up with four medium datasets. The small and medium datasets are updated dynamically at the beginning of each optimisation step and hence their composition might vary across time.

We use variable selection techniques minimising the AIC, BIC and HQ information criteria, employing several heuristic optimisation algorithms to handle the large set of regressors, including Simulated Annealing, Genetic Algorithm, MC^3 and Sequential testing. Furthermore, we have considered standard variable reduction methods, including Principal Components, Partial Least Squares and Bayesian Shrinkage regression. Our empirical analysis is focused on forecasting key economic indicators for the Euro Area, specifically: the quarterly GDP and consumption growth rates, and the monthly industrial production and HICP growth rate.

Our main findings suggest that the use of medium and small datasets indeed improves the forecasting ability of the models. Moreover, overall variable selection seems to perform better than reduction in our application. Among the various variable selection methods, there is no clear winner, so it may be worth implementing and comparing a few of them in empirical applications. Furthermore, we show that the use of out-of-sample forecasting using variable selection and variable reduction methodologies might be of importance to the applied research that focuses on the construction of (Cyclical) Composite Indicators.

Annex

6.A Tables

Table 6.A.1: Forecasting the GDP growth using 3MA transformation with large and AVGDYN small datasets.

	GDP Growth using 3MA							
	Large				AVGDYN			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	2.412	1.513	1.167	1.642	1.007	0.917	0.639	0.661
GA_{AIC}	4.102	1.200	1.698	1.859	1.111	0.909	0.649	0.655
MC^3_{AIC}	0.971	0.889	0.809	0.868	1.174	0.885	0.609	0.640
ST^1_{AIC}	1.126	0.894	0.826	0.768	1.083	0.829	0.720	0.682
ST^5_{AIC}	1.559	0.906	0.789	0.968	1.246	0.913	0.774	0.720
SA_{BIC}	1.365	0.836	0.837	1.006	1.084	0.810	0.652	0.625
GA_{BIC}	0.938	0.774	0.836	1.185	1.081	0.810	0.644	0.601
MC^3_{BIC}	1.079	0.722	0.781	0.593	1.084	0.810	0.652	0.625
SA_{HQ}	1.741	1.179	0.800	1.421	1.042	0.863	0.641	0.623
GA_{HQ}	1.264	0.784	0.826	0.778	0.955	0.735	0.630	0.608
MC^3_{HQ}	1.070	0.828	0.993	0.752	0.999	0.857	0.635	0.621
$PC\,(1)$	1.992	1.422	1.205	1.144	1.669	1.385	1.189	1.139
$PC\,(3)$	2.038	1.561	1.336	1.374	1.704	1.497	1.307	1.306
$PLS\,(1)$	1.830	1.352	1.175	1.175	1.604	1.357	1.179	1.168
$PLS\,(3)$	1.594	1.412	1.269	1.338	1.559	1.440	1.315	1.353
$BR\,(0.5N)$	2.827	2.017	1.876	1.959	2.009	1.746	1.566	1.557
$BR\,(2N)$	2.217	1.608	1.432	1.425	1.800	1.541	1.352	1.331
	Cross-Validation: 16 periods							

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.2: Forecasting the GDP growth using 3MA transformation with $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

	GDP Growth using 3MA							
	$\widehat{\alpha}_1$				$\widehat{\alpha}_2$			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	2.474	1.509	1.452	1.347	1.673	1.292	1.056	1.111
GA_{AIC}	1.779	0.902	1.139	1.131	3.161	1.895	3.792	4.717
MC^3_{AIC}	0.755	0.601	0.782	0.869	1.108	0.811	0.642	0.918
ST^1_{AIC}	0.763	0.576	0.570	0.551	1.025	2.123	0.601	0.624
ST^5_{AIC}	2.540	0.898	1.321	0.957	1.929	0.985	0.771	0.774
SA_{BIC}	0.816	0.690	1.028	0.914	1.100	0.678	0.887	0.970
GA_{BIC}	2.925	0.985	0.832	1.490	1.154	0.860	0.772	0.725
MC^3_{BIC}	0.799	0.919	0.700	0.727	0.984	0.505	0.680	0.686
SA_{HQ}	5.599	39.380	4.143	0.967	1.291	1.637	2.187	1.228
GA_{HQ}	1.806	3.268	1.306	1.879	1.353	1.149	1.439	1.023
MC^3_{HQ}	1.207	0.671	0.737	0.878	1.192	0.717	0.832	1.008
$PC(1)$	1.974	1.415	1.201	1.144	1.975	1.415	1.201	1.144
$PC(3)$	2.073	1.560	1.349	1.360	2.071	1.559	1.349	1.360
$PLS(1)$	1.912	1.391	1.194	1.172	1.916	1.392	1.195	1.172
$PLS(3)$	1.711	1.451	1.263	1.272	1.698	1.444	1.258	1.271
$BR(0.5N)$	2.674	1.960	1.768	1.760	2.676	1.957	1.775	1.760
$BR(2N)$	2.211	1.607	1.412	1.372	2.214	1.608	1.416	1.374
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_1$: 0.777, Avg. $\widehat{\alpha}_2$: 0.763								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.3: Forecasting the GDP growth using 3MObs transformation with $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

	GDP Growth using 3MA							
	$\widehat{\alpha}_3$				$\widehat{\alpha}_4$			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	1.973	1.141	1.197	0.909	1.994	1.962	1.578	2.360
GA_{AIC}	1.660	1.020	1.905	1.677	3.026	6.639	3.507	3.552
MC^3_{AIC}	0.912	0.663	0.698	0.666	1.239	1.112	0.961	0.642
ST^1_{AIC}	1.039	2.110	0.585	0.451	1.023	0.824	0.881	0.742
ST^5_{AIC}	1.848	0.938	0.681	0.780	1.248	0.657	1.062	0.558
SA_{BIC}	0.842	0.554	0.648	0.994	3.322	1.588	1.324	1.063
GA_{BIC}	0.805	0.832	0.631	0.659	0.676	0.769	0.695	0.692
MC^3_{BIC}	1.035	0.773	0.676	0.641	1.311	1.052	0.914	0.820
SA_{HQ}	1.742	1.053	0.913	1.823	2.303	3.368	1.206	2.306
GA_{HQ}	2.093	3.959	0.990	0.987	1.465	0.865	0.983	2.290
MC^3_{HQ}	0.890	0.818	0.696	0.840	1.301	1.011	0.735	0.590
$PC\,(1)$	1.975	1.415	1.201	1.145	1.985	1.419	1.204	1.144
$PC\,(3)$	2.072	1.560	1.347	1.359	2.061	1.561	1.341	1.364
$PLS\,(1)$	1.917	1.392	1.195	1.172	1.849	1.363	1.180	1.173
$PLS\,(3)$	1.693	1.442	1.256	1.272	1.545	1.405	1.259	1.317
$BR\,(0.5N)$	2.667	1.959	1.775	1.762	2.789	2.015	1.862	1.925
$BR\,(2N)$	2.212	1.608	1.416	1.375	2.204	1.609	1.428	1.412
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_3$: 0.761, Avg. $\widehat{\alpha}_4$: 0.915								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.4: Forecasting the GDP growth using 3MObs transformation with large and AVGDYN small datasets.

	GDP Growth using 3MObs							
	Large				AVGDYN			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	2.436	1.662	2.343	2.114	1.079	0.823	0.684	0.730
GA_{AIC}	1.626	0.944	1.056	1.104	1.027	0.809	0.664	0.726
MC^3_{AIC}	1.186	1.059	1.207	0.733	1.013	0.760	0.668	0.744
ST^1_{AIC}	0.759	1.022	0.988	0.736	1.019	0.760	0.679	0.666
ST^5_{AIC}	0.749	5.677	0.990	0.937	1.123	0.871	0.729	0.767
SA_{BIC}	1.681	0.684	0.695	0.943	1.147	0.821	0.673	0.689
GA_{BIC}	0.837	0.880	0.613	0.723	1.108	0.834	0.695	0.744
MC^3_{BIC}	1.103	0.718	0.727	0.827	1.219	0.898	0.711	0.703
SA_{HQ}	1.985	1.564	1.709	5.290	1.045	0.783	0.694	0.744
GA_{HQ}	1.049	0.988	0.668	1.511	1.094	0.794	0.704	0.717
MC^3_{HQ}	1.689	1.005	0.817	0.823	1.097	0.813	0.675	0.725
$PC\,(1)$	2.098	1.466	1.229	1.158	1.746	1.435	1.219	1.151
$PC\,(3)$	1.945	1.555	1.308	1.263	1.690	1.475	1.270	1.264
$PLS\,(1)$	1.731	1.296	1.145	1.133	1.571	1.340	1.184	1.171
$PLS\,(3)$	1.602	1.346	1.241	1.278	1.392	1.314	1.306	1.420
$BR\,(0.5N)$	2.681	2.169	1.981	2.043	1.726	1.670	1.685	1.842
$BR\,(2N)$	2.140	1.658	1.432	1.427	1.800	1.541	1.352	1.331
	Cross-Validation: 16 periods							

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.5: Forecasting the GDP growth using 3MObs transformation with $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

| | GDP Growth using 3MObs | | | | | | | |
| | $\widehat{\alpha}_1$ | | | | $\widehat{\alpha}_2$ | | | |
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	3.638	4.262	3.392	3.007	1.657	1.491	1.100	0.980
GA_{AIC}	1.435	1.611	1.051	2.670	1.986	2.586	0.810	1.518
MC^3_{AIC}	0.876	0.690	0.927	0.648	1.185	0.816	0.855	0.631
ST^1_{AIC}	0.888	0.766	0.654	0.772	1.044	0.799	0.711	0.709
ST^5_{AIC}	1.122	0.768	0.706	0.728	1.080	1.113	0.719	0.788
SA_{BIC}	0.699	0.865	0.689	0.703	0.892	0.846	0.866	0.951
GA_{BIC}	0.809	0.748	0.715	0.839	1.045	0.911	0.702	0.660
MC^3_{BIC}	0.860	0.813	0.805	0.773	1.182	0.680	0.805	0.724
SA_{HQ}	1.031	0.860	0.948	1.855	1.236	1.075	0.746	1.041
GA_{HQ}	1.208	0.803	0.682	0.632	1.401	0.952	1.021	1.401
MC^3_{HQ}	1.274	0.968	0.801	0.734	0.844	0.827	0.758	0.679
$PC\,(1)$	2.081	1.457	1.226	1.156	2.076	1.454	1.225	1.154
$PC\,(3)$	2.028	1.523	1.274	1.236	2.084	1.509	1.245	1.210
$PLS\,(1)$	1.833	1.337	1.158	1.133	1.865	1.352	1.162	1.132
$PLS\,(3)$	1.638	1.350	1.224	1.247	1.662	1.354	1.213	1.232
$BR\,(0.5N)$	2.500	2.075	1.860	1.873	2.461	1.990	1.792	1.783
$BR\,(2N)$	2.109	1.640	1.406	1.381	2.110	1.618	1.388	1.354
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_1$: 0.739, Avg. $\widehat{\alpha}_2$: 0.664								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.6: Forecasting the GDP growth using 3MObs transformation with $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

| | GDP Growth using 3MObs | | | | | | | |
| | $\widehat{\alpha}_3$ | | | | $\widehat{\alpha}_4$ | | | |
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.004	0.006	0.007	0.007	0.004	0.006	0.007	0.007
SA_{AIC}	2.547	2.642	1.492	1.548	1.890	1.089	1.503	1.722
GA_{AIC}	1.423	1.091	1.042	1.110	2.180	3.694	1.424	1.716
MC^3_{AIC}	0.975	1.043	0.726	0.697	1.081	0.958	0.972	0.971
ST^1_{AIC}	1.214	0.880	0.639	0.614	1.333	0.865	0.781	0.791
ST^5_{AIC}	0.978	0.902	1.015	1.116	1.069	0.936	1.030	1.296
SA_{BIC}	0.844	0.854	0.706	0.906	0.831	0.699	0.949	0.696
GA_{BIC}	1.086	0.958	0.721	0.694	0.844	0.695	0.574	0.615
MC^3_{BIC}	0.978	0.827	0.607	0.633	1.059	0.806	0.654	0.621
SA_{HQ}	1.031	1.164	0.906	0.885	1.757	1.281	1.075	1.118
GA_{HQ}	1.107	0.961	0.671	0.955	1.529	1.839	5.253	0.925
MC^3_{HQ}	0.849	0.803	0.678	0.733	1.068	0.792	0.806	0.787
$PC\,(1)$	2.075	1.452	1.223	1.154	2.075	1.454	1.225	1.155
$PC\,(3)$	2.107	1.485	1.214	1.180	2.089	1.501	1.236	1.201
$PLS\,(1)$	1.890	1.362	1.166	1.131	1.869	1.354	1.163	1.132
$PLS\,(3)$	1.697	1.354	1.204	1.220	1.673	1.354	1.212	1.228
$BR\,(0.5N)$	2.452	1.929	1.728	1.722	2.455	2.005	1.778	1.776
$BR\,(2N)$	2.112	1.598	1.367	1.337	2.107	1.623	1.383	1.353
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_3$: 0.604, Avg. $\widehat{\alpha}_4$: 0.652								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.7: Forecasting the Consumption growth using 3MA transformation with large and AVGDYN small datasets.

| | Consumption Growth using 3MA | | | | | | | |
| | Large | | | | AVGDYN | | | |
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	1.272	1.717	1.988	3.170	0.381	0.666	0.733	0.755
GA_{AIC}	1.031	3.026	1.384	28.594	0.381	0.701	0.752	0.779
MC^3_{AIC}	0.647	1.634	1.146	0.773	0.384	0.713	0.644	0.727
ST^1_{AIC}	0.374	0.861	0.711	0.762	0.375	0.643	0.616	0.659
ST^5_{AIC}	1.483	2.112	5.963	3.425	0.430	0.645	0.816	0.709
SA_{BIC}	0.620	0.913	0.815	0.541	0.376	0.757	0.827	0.782
GA_{BIC}	0.510	2.591	1.034	0.497	0.538	0.945	0.876	0.878
MC^3_{BIC}	0.577	0.850	0.770	0.710	0.425	0.827	0.919	0.796
SA_{HQ}	1.231	1.237	1.951	3.187	0.397	0.721	0.783	0.759
GA_{HQ}	0.640	1.166	1.221	1.732	0.372	0.721	0.739	0.751
MC^3_{HQ}	0.895	1.379	1.057	1.441	0.374	0.756	0.856	0.776
$PC\,(1)$	0.786	1.209	1.010	1.049	0.801	1.201	1.032	1.083
$PC\,(3)$	0.869	1.233	1.094	1.246	0.845	1.246	1.107	1.207
$PLS\,(1)$	0.751	1.144	0.956	1.034	0.761	1.155	0.979	1.069
$PLS\,(3)$	0.823	1.166	1.059	1.092	0.918	1.200	1.075	1.192
$BR\,(0.5N)$	1.421	2.428	2.028	2.367	1.093	1.864	1.610	1.852
$BR\,(2N)$	0.976	1.577	1.331	1.487	1.800	1.541	1.352	1.331
Cross-Validation: 16 periods								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.8: Forecasting the Consumption growth using 3MA transformation with $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

	Consumption Growth using 3MA							
	$\widehat{\alpha}_1$				$\widehat{\alpha}_2$			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	0.425	1.501	1.980	2.200	1.935	3.426	4.529	2.690
GA_{AIC}	2.374	5.654	6.310	1.632	1.298	2.809	2.687	2.679
MC^3_{AIC}	0.379	0.858	0.900	0.540	0.566	0.979	0.970	0.908
ST^1_{AIC}	0.477	0.736	1.034	2.930	0.407	0.894	0.879	0.766
ST^5_{AIC}	0.475	0.954	0.681	2.081	0.826	1.113	0.932	1.247
SA_{BIC}	0.568	1.162	1.053	0.822	0.492	0.832	0.777	0.921
GA_{BIC}	0.405	0.738	0.716	0.611	0.471	0.679	0.725	0.590
MC^3_{BIC}	0.417	1.001	0.829	0.899	0.476	0.832	0.619	0.629
SA_{HQ}	0.559	1.133	1.680	1.947	1.015	3.6484	1.650	1.167
GA_{HQ}	1.816	2.058	2.116	3.289	1.403	1.366	3.745	5.620
MC^3_{HQ}	0.564	0.915	0.991	1.017	0.470	1.335	0.796	0.911
$PC\,(1)$	0.794	1.217	1.021	1.058	0.794	1.217	1.021	1.058
$PC\,(3)$	0.867	1.225	1.090	1.241	0.867	1.224	1.092	1.241
$PLS\,(1)$	0.750	1.147	0.953	1.036	0.750	1.146	0.953	1.036
$PLS\,(3)$	0.821	1.160	1.073	1.118	0.823	1.158	1.073	1.118
$BR\,(0.5N)$	1.249	2.171	1.833	2.078	1.242	2.164	1.822	2.081
$BR\,(2N)$	0.934	1.504	1.285	1.416	0.932	1.502	1.280	1.420
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_1$: 0.777, Avg. $\widehat{\alpha}_2$: 0.763								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.9: Forecasting the Consumption growth using 3MA transformation with $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

| | Consumption Growth using 3MA | | | | | | | |
| | $\widehat{\alpha}_3$ | | | | $\widehat{\alpha}_4$ | | | |
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	1.234	0.786	1.830	2.197	1.041	1.009	1.092	2.051
GA_{AIC}	4.394	7.329	7.600	3.729	1.193	2.774	1.792	1.117
MC^3_{AIC}	0.406	1.142	0.904	0.720	0.453	1.164	0.935	0.779
ST^1_{AIC}	0.414	0.805	0.896	0.612	0.527	1.199	1.009	0.741
ST^5_{AIC}	1.990	5.315	2.045	1.655	1.089	1.283	0.938	2.017
SA_{BIC}	0.790	1.111	0.942	2.782	1.112	1.217	1.966	0.640
GA_{BIC}	0.607	0.993	0.849	0.967	0.547	1.766	0.839	1.196
MC^3_{BIC}	0.803	1.225	0.787	1.053	0.639	0.776	0.947	0.766
SA_{HQ}	2.420	1.581	3.297	4.393	2.388	2.121	6.280	4.595
GA_{HQ}	0.796	1.357	1.908	1.006	1.281	2.105	1.171	2.192
MC^3_{HQ}	0.425	1.062	1.045	0.641	0.541	1.129	0.978	0.980
$PC\,(1)$	0.794	1.217	1.021	1.057	0.789	1.212	1.014	1.052
$PC\,(3)$	0.867	1.224	1.092	1.241	0.865	1.227	1.091	1.238
$PLS\,(1)$	0.750	1.147	0.954	1.036	0.750	1.144	0.955	1.036
$PLS\,(3)$	0.822	1.157	1.073	1.118	0.820	1.163	1.064	1.105
$BR\,(0.5N)$	1.239	2.168	1.825	2.083	1.391	2.401	2.011	2.323
$BR\,(2N)$	0.931	1.503	1.280	1.420	0.968	1.569	1.328	1.474
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_3$: 0.761, Avg. $\widehat{\alpha}_4$: 0.915								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MA denotes the 3-Months Averaging transformation method. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.10: Forecasting the Consumption growth using 3MObs transformation with large and AVGDYN small datasets.

	Consumption Growth using 3MObs							
	Large				AVGDYN			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	1.641	2.589	3.300	1.872	0.333	0.783	0.572	0.864
GA_{AIC}	0.983	1.319	1.600	2.236	0.530	0.854	0.708	0.892
MC^3_{AIC}	0.481	0.768	0.915	1.089	0.707	0.963	0.787	0.974
ST^1_{AIC}	0.430	0.765	0.928	0.688	0.340	0.663	0.614	0.666
ST^5_{AIC}	0.617	1.282	1.102	1.106	0.369	0.617	0.624	0.605
SA_{BIC}	0.488	1.489	1.165	0.896	0.549	0.783	0.814	0.823
GA_{BIC}	0.533	0.911	0.679	0.814	0.594	0.814	0.835	0.826
MC^3_{BIC}	0.386	1.187	1.007	0.892	0.558	0.850	0.824	0.804
SA_{HQ}	1.483	2.122	3.100	2.069	0.691	0.903	0.906	0.888
GA_{HQ}	0.668	0.826	1.329	1.393	0.791	1.165	0.929	1.013
MC^3_{HQ}	0.541	1.250	0.912	0.922	0.674	0.813	0.867	0.968
$PC\,(1)$	0.763	1.188	0.970	1.028	0.780	1.181	0.994	1.056
$PC\,(3)$	0.815	1.282	1.115	1.378	0.841	1.230	1.153	1.184
$PLS\,(1)$	0.744	1.140	0.944	1.034	0.760	1.141	0.960	1.051
$PLS\,(3)$	0.809	1.211	1.156	1.200	0.821	1.221	1.053	1.130
$BR\,(0.5N)$	1.377	2.570	2.219	2.514	1.072	1.968	1.713	2.141
$BR\,(2N)$	0.948	1.626	1.364	1.518	1.800	1.541	1.352	1.331
	Cross-Validation: 16 periods							

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.11: Forecasting the Consumption growth using 3MObs transformation with $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

| | Consumption Growth using 3MObs | | | | | | | |
| | $\widehat{\alpha}_1$ | | | | $\widehat{\alpha}_2$ | | | |
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	1.389	3.348	5.016	1.750	1.203	2.092	1.969	2.472
GA_{AIC}	1.202	1.234	0.689	1.551	1.422	0.966	2.493	0.954
MC^3_{AIC}	0.475	0.767	0.911	1.160	0.522	0.983	0.663	0.789
ST^1_{AIC}	1.521	0.726	7.505	0.820	0.426	0.786	0.700	0.620
ST^5_{AIC}	0.886	1.255	1.274	7.625	0.483	1.414	0.905	2.678
SA_{BIC}	0.468	1.845	0.627	1.015	0.424	0.807	0.527	0.538
GA_{BIC}	0.465	0.594	0.636	0.725	0.502	0.879	0.630	0.796
MC^3_{BIC}	0.467	0.784	0.654	1.402	0.513	0.862	0.594	0.721
SA_{HQ}	0.938	3.930	6.705	0.866	0.848	0.907	1.251	1.154
GA_{HQ}	1.207	1.709	1.905	1.415	0.655	0.865	0.797	0.717
MC^3_{HQ}	0.401	1.105	0.672	0.980	0.489	0.716	0.683	0.463
$PC\,(1)$	0.769	1.202	0.979	1.037	0.770	1.204	0.981	1.038
$PC\,(3)$	0.832	1.287	1.133	1.317	0.838	1.283	1.168	1.291
$PLS\,(1)$	0.742	1.138	0.944	1.030	0.739	1.132	0.943	1.029
$PLS\,(3)$	0.846	1.244	1.148	1.176	0.857	1.249	1.155	1.182
$BR\,(0.5N)$	1.240	2.312	2.045	2.234	1.198	2.161	1.939	2.133
$BR\,(2N)$	0.922	1.554	1.330	1.455	0.914	1.511	1.309	1.423
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_1$: 0.739, Avg. $\widehat{\alpha}_2$: 0.664								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.12: Forecasting the Consumption growth using 3MObs transformation with $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

	Consumption Growth using 3MObs							
	$\widehat{\alpha}_3$				$\widehat{\alpha}_4$			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
AR(1)	0.014	0.009	0.010	0.010	0.014	0.009	0.010	0.010
SA_{AIC}	1.340	1.949	1.311	0.987	2.119	2.693	1.755	2.084
GA_{AIC}	0.830	1.111	0.929	1.863	0.514	0.829	2.498	1.440
MC^3_{AIC}	0.529	1.245	0.720	0.659	0.574	0.505	0.658	0.689
ST^1_{AIC}	0.427	0.602	0.635	1.817	0.424	0.726	0.626	0.529
ST^5_{AIC}	1.832	0.908	2.424	1.324	0.782	2.484	1.769	9.109
SA_{BIC}	0.357	0.894	0.737	0.582	0.446	0.628	0.568	0.596
GA_{BIC}	0.437	0.671	0.716	0.465	0.537	0.604	0.563	0.569
MC^3_{BIC}	0.457	0.775	0.769	0.495	0.440	0.747	0.898	0.635
SA_{HQ}	0.512	1.122	0.836	1.334	0.913	0.618	1.249	1.392
GA_{HQ}	0.735	0.907	0.874	1.846	0.759	1.214	0.824	2.176
MC^3_{HQ}	0.536	0.946	0.838	0.761	0.676	0.686	0.691	1.594
$PC\,(1)$	0.771	1.205	0.983	1.039	0.771	1.204	0.982	1.038
$PC\,(3)$	0.835	1.271	1.168	1.252	0.835	1.283	1.165	1.270
$PLS\,(1)$	0.740	1.135	0.946	1.023	0.738	1.132	0.943	1.027
$PLS\,(3)$	0.853	1.234	1.167	1.163	0.856	1.249	1.163	1.177
$BR\,(0.5N)$	1.158	2.070	1.839	2.051	1.198	2.146	1.918	2.130
$BR\,(2N)$	0.904	1.486	1.290	1.400	0.913	1.509	1.305	1.422
Cross-Validation: 16 periods, Avg. $\widehat{\alpha}_3$: 0.604, Avg. $\widehat{\alpha}_4$: 0.652								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. 3MObs denotes the 3rd month of each quarter transformation method. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.13: Forecasting the Industrial Production growth using large and AVGDYN small datasets.

| | Industrial Production Growth | | | | | | | |
| | Large | | | | AVGDYN | | | |
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.007	0.008	0.008	0.011	0.007	0.008	0.008	0.011
SA_{AIC}	2.250	2.065	7.844	5.991	1.050	0.917	1.024	1.014
GA_{AIC}	1.307	1.457	1.545	1.613	1.006	0.940	1.029	1.025
MC^3_{AIC}	1.117	1.149	1.218	1.091	1.033	0.924	1.019	1.011
ST^1_{AIC}	1.037	1.022	0.973	0.995	0.972	0.930	1.012	1.034
ST^5_{AIC}	1.391	1.204	1.641	1.515	0.980	0.893	0.993	1.061
SA_{BIC}	1.076	1.001	1.050	1.057	0.980	0.927	1.001	1.002
GA_{BIC}	1.051	1.017	1.086	1.040	0.983	0.932	1.000	1.001
MC^3_{BIC}	1.001	0.964	1.003	1.023	0.982	0.929	1.005	1.003
SA_{HQ}	1.489	1.302	2.497	2.713	0.977	0.920	1.010	1.013
GA_{HQ}	0.983	1.053	1.086	1.171	0.992	0.924	1.020	1.014
MC^3_{HQ}	1.080	1.049	1.055	0.997	0.984	0.927	1.006	1.007
$PC(1)$	1.055	0.982	0.999	0.983	1.052	0.982	1.000	0.984
$PC(3)$	1.015	0.908	1.016	0.975	1.068	0.926	1.028	0.974
$PLS(1)$	1.028	0.929	1.012	0.999	1.039	0.949	1.015	0.995
$PLS(3)$	0.996	0.918	1.009	0.979	0.974	0.939	1.016	0.985
$BR(0.5N)$	1.101	1.045	1.260	1.002	1.005	0.962	1.132	1.047
$BR(2N)$	1.010	0.951	1.079	0.963	0.993	0.938	1.057	1.004
	Cross-Validation: 86 periods							

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.14: Forecasting the Industrial Production growth using $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

	Industrial Production Growth							
	$\widehat{\alpha}_1$				$\widehat{\alpha}_2$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.007	0.008	0.008	0.011	0.007	0.008	0.008	0.011
SA_{AIC}	2.094	1.449	2.763	3.043	1.978	1.362	3.483	3.526
GA_{AIC}	1.371	1.256	1.541	1.561	1.387	1.305	1.802	1.441
MC^3_{AIC}	1.039	1.015	1.124	1.044	0.977	0.990	1.106	1.063
ST^1_{AIC}	1.070	0.970	1.050	1.061	1.090	1.163	1.234	2.338
ST^5_{AIC}	1.647	1.450	1.957	4.240	2.652	5.659	8.869	2.352
SA_{BIC}	0.979	1.006	1.101	0.973	1.054	0.977	1.055	0.986
GA_{BIC}	0.985	0.939	1.122	1.037	1.054	0.994	1.124	1.033
MC^3_{BIC}	0.979	0.963	1.096	0.977	0.988	0.954	1.036	1.004
SA_{HQ}	1.260	1.328	4.621	1.196	1.164	1.119	1.520	1.387
GA_{HQ}	1.249	1.108	1.401	1.430	1.245	1.036	1.264	1.507
MC^3_{HQ}	1.030	1.021	1.154	1.043	1.006	1.006	1.190	1.009
$PC(1)$	1.055	0.982	0.999	0.982	1.055	0.982	0.999	0.983
$PC(3)$	1.009	0.910	1.017	0.980	1.009	0.911	1.018	0.981
$PLS(1)$	1.031	0.927	1.012	0.999	1.031	0.926	1.012	0.999
$PLS(3)$	0.990	0.912	1.010	0.985	0.990	0.911	1.010	0.985
$BR(0.5N)$	1.067	1.029	1.220	1.014	1.050	1.033	1.213	1.014
$BR(2N)$	0.994	0.940	1.069	0.981	0.987	0.941	1.065	0.981
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_1$: 0.763, Avg. $\widehat{\alpha}_2$: 0.738								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.15: Forecasting the Industrial Production growth using $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

	Industrial Production Growth							
	$\widehat{\alpha}_3$				$\widehat{\alpha}_4$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.007	0.008	0.008	0.011	0.007	0.008	0.008	0.011
SA_{AIC}	1.755	2.204	2.579	3.009	2.750	8.032	6.825	9.921
GA_{AIC}	1.215	1.373	1.888	1.360	1.331	1.436	1.460	2.632
MC^3_{AIC}	0.992	0.963	1.150	1.053	0.997	1.081	1.098	1.106
ST^1_{AIC}	1.111	0.962	1.202	1.184	1.110	0.987	1.051	1.091
ST^5_{AIC}	2.877	2.376	3.387	2.721	1.535	2.029	3.094	2.626
SA_{BIC}	1.035	1.025	1.103	1.013	1.059	0.984	1.108	1.041
GA_{BIC}	0.978	0.977	1.049	0.999	1.080	1.047	1.136	1.163
MC^3_{BIC}	1.007	0.950	1.065	0.980	1.003	0.955	1.045	0.986
SA_{HQ}	1.230	1.144	1.563	1.495	2.564	1.637	2.079	5.730
GA_{HQ}	1.223	1.072	1.427	1.515	1.179	1.074	1.282	1.246
MC^3_{HQ}	1.000	1.006	1.112	0.987	0.944	1.019	1.107	1.021
$PC\,(1)$	1.055	0.982	0.999	0.983	1.055	0.982	0.999	0.983
$PC\,(3)$	1.009	0.911	1.017	0.982	1.011	0.911	1.018	0.978
$PLS\,(1)$	1.032	0.926	1.012	0.999	1.029	0.928	1.012	0.999
$PLS\,(3)$	0.993	0.911	1.010	0.985	0.995	0.915	1.009	0.982
$BR\,(0.5N)$	1.045	1.035	1.223	1.014	1.086	1.043	1.241	1.022
$BR\,(2N)$	0.986	0.942	1.070	0.980	1.005	0.946	1.075	0.976
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_3$: 0.732, Avg. $\widehat{\alpha}_4$: 0.906								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.16: Forecasting the HICP growth using large and AVGDYN small datasets.

	HICP Growth							
	Large				AVGDYN			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.474	3.494	1.745	1.720	0.868	0.869	0.873	0.921
GA_{AIC}	1.072	1.382	4.091	2.123	0.873	0.863	0.869	0.924
MC^3_{AIC}	1.040	0.993	1.088	0.786	0.871	0.847	0.870	0.928
ST^1_{AIC}	0.815	0.887	0.934	0.885	0.851	0.840	0.847	0.888
ST^5_{AIC}	0.961	1.443	2.407	1.385	0.880	0.862	0.846	0.913
SA_{BIC}	0.897	0.924	0.949	0.822	0.864	0.857	0.872	0.911
GA_{BIC}	0.875	0.956	0.918	0.865	0.864	0.857	0.872	0.911
MC^3_{BIC}	0.869	0.941	0.867	0.805	0.880	0.867	0.867	0.917
SA_{HQ}	1.234	1.632	1.418	1.916	0.858	0.849	0.871	0.921
GA_{HQ}	0.876	1.021	0.972	0.776	0.858	0.849	0.871	0.921
MC^3_{HQ}	0.847	0.908	0.960	0.797	0.851	0.867	0.877	0.912
$PC\,(1)$	1.015	1.016	1.011	1.031	1.013	1.017	1.012	1.033
$PC\,(3)$	0.969	1.013	0.993	1.058	1.007	1.009	1.005	1.063
$PLS\,(1)$	1.001	0.992	0.971	1.004	1.011	1.013	0.986	1.008
$PLS\,(3)$	0.994	0.972	0.980	1.010	1.021	0.995	1.001	1.036
$BR\,(0.5N)$	1.171	1.264	1.412	1.623	1.058	1.087	1.122	1.390
$BR\,(2N)$	1.039	1.081	1.133	1.232	1.031	1.035	1.061	1.197
Cross-Validation: 86 periods								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.17: Forecasting the HICP growth using $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

	HICP Growth							
	$\widehat{\alpha}_1$				$\widehat{\alpha}_2$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.012	6.471	1.508	4.390	1.230	2.278	2.341	5.663
GA_{AIC}	1.004	1.336	1.350	1.305	0.938	1.120	1.081	0.863
MC^3_{AIC}	0.925	0.939	0.941	0.824	0.841	0.914	0.912	0.815
ST^1_{AIC}	1.048	1.152	0.928	1.374	0.881	1.697	5.041	1.058
ST^5_{AIC}	2.503	3.330	1.972	4.269	2.014	2.346	2.021	3.509
SA_{BIC}	0.863	0.883	0.873	0.855	0.862	0.891	0.870	0.861
GA_{BIC}	0.836	0.897	0.879	0.886	0.875	0.907	0.877	0.833
MC^3_{BIC}	0.845	0.915	0.893	0.854	0.855	0.899	0.863	0.834
SA_{HQ}	0.952	1.077	1.086	1.084	0.909	1.209	0.963	0.989
GA_{HQ}	0.873	0.948	1.023	0.858	0.921	0.957	0.933	0.793
MC^3_{HQ}	0.845	0.917	0.866	0.782	0.875	0.932	0.877	0.803
$PC\,(1)$	1.014	1.015	1.010	1.030	1.014	1.015	1.010	1.030
$PC\,(3)$	0.964	1.008	0.982	1.049	0.965	1.007	0.981	1.047
$PLS\,(1)$	0.995	0.991	0.964	0.999	0.996	0.991	0.964	0.997
$PLS\,(3)$	1.005	0.994	0.969	1.015	1.005	0.995	0.972	1.016
$BR\,(0.5N)$	1.121	1.237	1.345	1.546	1.117	1.235	1.336	1.541
$BR\,(2N)$	1.029	1.081	1.108	1.202	1.027	1.081	1.107	1.200
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_1$: 0.763, Avg. $\widehat{\alpha}_2$: 0.738								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.18: Forecasting the HICP growth using $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

	HICP Growth							
	$\widehat{\alpha}_3$				$\widehat{\alpha}_4$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.141	2.851	1.511	1.964	1.364	5.506	5.952	2.269
GA_{AIC}	0.933	1.139	1.111	1.361	1.063	1.088	1.254	1.150
MC^3_{AIC}	0.903	0.917	0.935	0.799	0.935	0.986	0.920	0.890
ST^1_{AIC}	0.908	0.925	0.906	0.841	0.891	0.888	0.911	0.854
ST^5_{AIC}	1.228	1.596	2.051	3.208	1.673	3.088	2.328	2.110
SA_{BIC}	0.867	0.979	0.969	0.847	0.875	0.894	0.893	0.796
GA_{BIC}	0.859	0.910	0.913	0.853	0.855	0.927	0.888	0.900
MC^3_{BIC}	0.894	0.918	0.904	0.815	0.881	0.922	0.903	0.839
SA_{HQ}	0.863	1.058	0.997	0.969	1.948	1.769	1.389	0.972
GA_{HQ}	0.951	1.041	0.970	0.799	0.948	1.047	0.920	0.810
MC^3_{HQ}	0.853	0.960	0.973	0.823	0.920	1.003	0.974	0.920
$PC\,(1)$	1.014	1.015	1.010	1.030	1.015	1.016	1.011	1.030
$PC\,(3)$	0.965	1.008	0.981	1.047	0.963	1.014	0.990	1.048
$PLS\,(1)$	0.996	0.991	0.964	0.997	0.996	0.990	0.968	1.000
$PLS\,(3)$	1.007	0.997	0.973	1.016	0.994	0.980	0.976	1.010
$BR\,(0.5N)$	1.125	1.239	1.333	1.530	1.146	1.260	1.392	1.576
$BR\,(2N)$	1.030	1.082	1.106	1.196	1.030	1.083	1.128	1.210
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_3$: 0.732, Avg. $\widehat{\alpha}_4$: 0.906								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

Table 6.A.19: Forecasting the HICP growth including AR(1) using large and AVGDYN small datasets.

	HICP Growth with AR(1)							
	Large				AVGDYN			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.470	1.861	4.378	2.235	0.883	0.877	0.857	0.925
GA_{AIC}	1.200	1.414	1.586	1.914	0.873	0.877	0.867	0.924
MC^3_{AIC}	0.894	1.008	0.915	0.722	0.858	0.878	0.873	0.946
ST^1_{AIC}	0.804	0.902	0.872	0.875	0.849	0.857	0.846	0.884
ST^5_{AIC}	0.905	1.235	1.262	1.000	0.877	0.856	0.851	0.961
SA_{BIC}	0.865	0.947	0.994	0.850	0.862	0.866	0.867	0.916
GA_{BIC}	0.854	0.924	0.863	0.901	0.862	0.866	0.867	0.916
MC^3_{BIC}	0.852	0.921	0.991	0.871	0.864	0.868	0.864	0.915
SA_{HQ}	1.062	2.630	1.238	0.975	0.853	0.860	0.880	0.940
GA_{HQ}	0.923	1.041	1.070	0.911	0.853	0.860	0.879	0.941
MC^3_{HQ}	0.920	0.941	0.965	0.779	0.855	0.857	0.893	0.944
$PC(1)$	1.016	1.017	1.011	1.033	1.008	1.018	1.013	1.035
$PC(3)$	0.973	1.019	0.994	1.093	0.996	1.007	1.016	1.076
$PLS(1)$	0.999	0.992	0.967	1.005	1.005	1.001	0.988	1.004
$PLS(3)$	0.987	0.971	0.975	1.028	1.010	0.999	1.003	1.042
$BR(0.5N)$	1.164	1.272	1.425	1.612	1.041	1.082	1.141	1.393
$BR(2N)$	1.035	1.081	1.138	1.228	1.020	1.034	1.074	1.199
	Cross-Validation: 86 periods							

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. AVGDYN selects one representative variable within each group of indicators.

Table 6.A.20: Forecasting the HICP growth including AR(1) using $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ medium datasets.

	HICP Growth with AR(1)							
	$\widehat{\alpha}_1$				$\widehat{\alpha}_2$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.216	1.709	2.337	4.536	1.480	2.258	2.918	2.397
GA_{AIC}	1.068	1.204	1.167	1.422	0.943	0.985	1.142	0.945
MC^3_{AIC}	0.881	0.950	0.957	0.941	0.881	0.966	0.935	0.832
ST^1_{AIC}	0.883	1.102	0.943	1.232	0.989	1.057	1.086	4.809
ST^5_{AIC}	1.658	5.523	3.822	2.096	1.262	7.465	1.774	2.616
SA_{BIC}	0.860	0.871	0.885	0.870	0.884	0.875	0.877	0.853
GA_{BIC}	0.860	0.899	0.894	0.843	0.879	0.910	0.915	0.865
MC^3_{BIC}	0.859	0.890	0.915	0.811	0.838	0.870	0.919	0.846
SA_{HQ}	1.126	2.440	4.556	1.239	0.933	1.015	1.112	0.903
GA_{HQ}	0.927	0.996	1.002	0.792	0.957	0.998	0.969	0.809
MC^3_{HQ}	0.878	0.866	0.892	0.835	0.903	0.920	0.921	0.787
$PC\,(1)$	1.009	1.016	1.010	1.031	1.008	1.016	1.010	1.031
$PC\,(3)$	0.960	1.014	0.984	1.059	0.958	1.013	0.983	1.056
$PLS\,(1)$	0.988	0.987	0.960	0.999	0.988	0.986	0.962	0.998
$PLS\,(3)$	0.987	0.988	0.963	1.020	0.986	0.989	0.966	1.015
$BR\,(0.5N)$	1.098	1.230	1.359	1.536	1.100	1.233	1.354	1.534
$BR\,(2N)$	1.017	1.078	1.116	1.197	1.016	1.079	1.114	1.197
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_1$: 0.765, Avg. $\widehat{\alpha}_2$: 0.741								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ denote the estimator of the cross-sectional exponent.

Table 6.A.21: Forecasting the HICP growth including AR(1) using $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ medium datasets.

	HICP Growth with AR(1)							
	$\widehat{\alpha}_3$				$\widehat{\alpha}_4$			
	h=1	h=3	h=6	h=12	h=1	h=3	h=6	h=12
AR(1)	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
SA_{AIC}	1.381	1.517	1.585	1.158	1.620	1.519	1.825	2.804
GA_{AIC}	1.102	1.146	1.284	1.294	0.996	1.265	1.422	2.653
MC^3_{AIC}	0.890	0.968	0.945	0.818	0.891	0.954	0.911	0.763
ST^1_{AIC}	0.899	1.060	6.273	1.706	1.152	1.158	1.282	0.884
ST^5_{AIC}	2.317	3.907	1.743	1.013	1.461	6.634	9.218	2.211
SA_{BIC}	0.856	0.932	0.871	0.848	0.858	0.937	0.857	0.790
GA_{BIC}	0.912	0.875	0.885	0.874	0.856	0.895	0.850	0.875
MC^3_{BIC}	0.888	0.872	0.936	0.830	0.846	0.924	0.929	0.868
SA_{HQ}	0.958	1.050	1.018	1.379	0.959	1.002	0.924	0.923
GA_{HQ}	0.878	0.941	1.002	0.983	0.884	0.963	0.983	0.935
MC^3_{HQ}	0.894	0.940	0.968	0.831	0.847	1.011	0.937	0.893
$PC\,(1)$	1.009	1.015	1.010	1.031	1.009	1.015	1.010	1.032
$PC\,(3)$	0.958	1.013	0.983	1.055	0.962	1.014	0.989	1.074
$PLS\,(1)$	0.988	0.986	0.962	0.997	0.991	0.988	0.966	1.001
$PLS\,(3)$	0.985	0.991	0.967	1.015	0.987	0.977	0.972	1.021
$BR\,(0.5N)$	1.106	1.236	1.349	1.535	1.113	1.248	1.404	1.564
$BR\,(2N)$	1.019	1.081	1.112	1.198	1.020	1.076	1.132	1.206
Cross-Validation: 86 periods, Avg. $\widehat{\alpha}_3$: 0.735, Avg. $\widehat{\alpha}_4$: 0.905								

Notes: h denotes the forecast steps ahead. $AR(1)$ is the benchmark model. SA, GA, MC^3 denote the Simulated Annealing, Genetic Algorithm and MC^3 algorithm respectively. ST denotes the sequential Testing at 1% and 5% significance levels. AIC, BIC and HQ denotes the criteria that are minimised by the optimisation algorithms. PC, PLS and BR denote the methods of Principal Components, Partial Least Squares and Bayesian Shrinkage Regression. The factors used are 1 and 3 and the shrinkage parameters are $v = 0.5N$ and $v = 2N$. $AR(1)$ presents the actual RMSFE. All other methods present their relative (to the benchmark) RMSFE. $\widehat{\alpha}_3$ and $\widehat{\alpha}_4$ denote the estimator of the cross-sectional exponent.

6.B Labels and Groups of Variables used and Transformations

Labels and Groups of Variables used

#	Group	Label	#	Group	Label
1	A	CP-HI00XEFU	100	K	1000-PERS-LM-UN-T-LE25
2	A	CP-HI00XEF	101	K	1000-PERS-LM-UN-T-TOT
3	A	CP-HI00XES	102	L	IS-IP
4	A	CP-HI00XE	103	L	IS-IP-F-CC1
5	A	CP-HI00XTB	104	L	IS-IP-F-CC2
6	A	CP-HI00	105	L	IS-IP-F
7	A	CP-HI01	106	M	B-D-IS-IP
8	A	CP-HI02	107	M	B-C-IS-IP
9	A	CP-HI03	108	M	B-IS-IP
10	A	CP-HI04	109	M	C-IS-IP
11	A	CP-HI05	110	M	C-ORD-IS-IP
12	A	CP-HI06	111	M	D-IS-IP
13	A	CP-HI07	112	M	MIG-CAG-IS-IP
14	A	CP-HI08	113	M	MIG-COG-IS-IP
15	A	CP-HI09	114	M	MIG-DCOG-IS-IP
16	A	CP-HI10	115	M	MIG-ING-IS-IP
17	A	CP-HI11	116	M	MIG-NDCOG-IS-IP
18	A	CP-HI12	117	N	IS-EPI-F
19	A	CP-HIE	118	N	B-D-IS-EPI
20	A	CP-HIF	119	N	B-E36-IS-EPI
21	B	B-D-IS-PPI	120	N	B-C-IS-EPI
22	B	B-E36-IS-PPI	121	N	B-IS-EPI
23	B	B-C-D-IS-PPI	122	N	B-TO-E36-IS-EPI
24	B	B-IS-PPI	123	N	C-IS-EPI
25	B	B-TO-E36-IS-PPI	124	N	D35-E36-IS-EPI
26	B	C-IS-PPI	125	N	D-IS-EPI
27	B	C-ORD-IS-PPI	126	N	E36-IS-EPI
28	B	D-IS-PPI	127	N	E-IS-EPI
29	B	E36-IS-PPI	128	N	MIG-CAG-IS-EPI
30	B	MIG-CAG-IS-PPI	129	N	MIG-COG-IS-EPI
31	B	MIG-COG-IS-PPI	130	N	MIG-DCOG-IS-EPI
32	B	MIG-DCOG-IS-PPI	131	N	MIG-ING-IS-EPI
33	B	MIG-ING-IS-PPI	132	N	MIG-NDCOG-IS-EPI
34	B	MIG-NDCOG-IS-PPI	133	N	MIG-NRG-IS-EPI
35	B	MIG-NRG-IS-PPI	134	N	G45-IS-EPI
36	B	D35-E36-IS-IMPR	135	O	B-C-IS-ITD
37	B	IS-WSI-F	136	O	B-C-IS-ITND
38	B	B-D-IS-WSI	137	O	B-C-IS-ITT
39	B	B-E36-IS-WSI	138	O	C-IS-ITD
40	B	B-C-D-IS-WSI	139	O	C-IS-ITND
41	B	B-IS-WSI	140	O	C-IS-ITT
42	B	B-TO-E36-IS-WSI	141	O	C-ORD-IS-ITD
43	B	C-IS-WSI	142	O	C-ORD-IS-ITND
44	B	D35-E36-IS-WSI	143	O	C-ORD-IS-ITT
45	B	D-IS-WSI	144	O	MIG-CAG-IS-ITD
46	B	E36-IS-WSI	145	O	MIG-CAG-IS-ITND
47	B	MIG-CAG-IS-WSI	146	O	MIG-CAG-IS-ITT
48	B	MIG-COG-IS-WSI	147	O	MIG-COG-IS-ITD
49	B	MIG-DCOG-IS-WSI	148	O	MIG-COG-IS-ITND
50	B	MIG-ING-IS-WSI	149	O	MIG-COG-IS-ITT
51	B	MIG-NDCOG-IS-WSI	150	O	MIG-DCOG-IS-ITD

Labels and Groups of Variables used

#	Group	Label	#	Group	Label
52	B	MIG-NRG-IS-WSI	151	O	MIG-DCOG-IS-ITND
53	C	BS-CCI-BAL	152	O	MIG-DCOG-IS-ITT
54	C	BS-CEME-BAL	153	O	MIG-ING-IS-ITD
55	C	BS-COB-BAL	154	O	MIG-ING-IS-ITND
56	C	BS-CPE-BAL	155	O	MIG-ING-IS-ITT
57	C	BS-CTA-BAL	156	O	MIG-NDCOG-IS-ITD
58	D	BS-BCI	157	O	MIG-NDCOG-IS-ITT
59	E	BS-CSMCI	158	P	IS-PEI-F-CC11-X-CC113
60	E	BS-FS-LY	159	P	IS-IPI-F-CC11-X-CC113
61	E	BS-FS-NY	160	P	IS-HWI-F
62	E	BS-GES-LY	161	P	C-ORD-IS-IO
63	E	BS-GES-NY	162	P	C-ORD-X-C30-IS-IO
64	E	BS-MP-NY	163	Q	IS-CAR
65	E	BS-MP-PR	164	R	FOOD-IS-DIT
66	E	BS-PT-LY	165	R	IS-DIT
67	E	BS-PT-NY	166	R	NFOOD-IS-DIT
68	E	BS-SFSH	167	R	NFOOD-X-G473-IS-DIT
69	E	BS-SV-NY	168	R	X-G473-IS-DIT
70	E	BS-SV-PR	169	S	M1
71	E	BS-UE-NY	170	S	M2
72	F	BS-ICI	171	S	M3
73	F	BS-IEME	172	T	3MI-RT
74	F	BS-IEOB	173	T	LTGBY-RT
75	F	BS-IOB	174	U	EXA-RT-USD
76	F	BS-IPE	175	U	EXA-RT-JPY
77	F	BS-IPT	176	U	EXA-RT-GBP
78	F	BS-ISFP	177	V	BDSHRPRCF
79	F	BS-ISPE	178	V	DJES50I
80	G	BS-RAS	179	V	EMECB2Y
81	G	BS-RCI	180	V	EMECB3Y
82	G	BS-REBS	181	V	EMECB5Y
83	G	BS-REM	182	V	EMECB7Y
84	G	BS-ROP	183	V	EMGBOND
85	G	BS-RPBS	184	V	FIBOR1Y
86	H	BS-SABC	185	V	FIBOR3M
87	H	BS-SAEM	186	V	FIBOR6M
88	H	BS-SARM	187	V	BDWU1032R
89	H	BS-SCI	188	V	BDWU0022R
90	H	BS-SERM	189	V	BDEBDBSIA
91	I	BS-CSMCI-B	190	V	BDECBXDGA
92	I	BS-ESI-I	191	V	BDECBXDMA
93	I	BS-ICI-BAL	192	V	BDECBXNGA
94	I	BS-RCI-BAL	193	V	BDECBXNOA
95	I	BS-SCI-BAL	194	V	BDECBXOLA
96	J	RT-LM-UN-T-GT25	195	V	BDECBXLIA
97	J	RT-LM-UN-T-LE25			
98	J	RT-LM-UN-T-TOT			
99	K	1000-PERS-LM-UN-T-GT25			

6.C Transformations

Transformations

#	Label	#	Label	#	Label	#	Label
1	First Diff., Logs	51	First Diff., Logs	101	No Change	151	First Diff., Logs
2	First Diff., Logs	52	First Diff., Logs	102	First Diff., Logs	152	First Diff., Logs
3	First Diff., Logs	53	No Change	103	First Diff., Logs	153	First Diff., Logs
4	First Diff., Logs	54	No Change	104	First Diff., Logs	154	First Diff., Logs
5	First Diff., Logs	55	No Change	105	First Diff., Logs	155	First Diff., Logs
6	First Diff., Logs	56	No Change	106	First Diff., Logs	156	First Diff., Logs
7	First Diff., Logs	57	No Change	107	First Diff., Logs	157	First Diff., Logs
8	First Diff., Logs	58	No Change	108	First Diff., Logs	158	First Diff., Logs
9	First Diff., Logs	59	No Change	109	First Diff., Logs	159	First Diff., Logs
10	First Diff., Logs	60	No Change	110	First Diff., Logs	160	First Diff., Logs
11	First Diff., Logs	61	No Change	111	First Diff., Logs	161	First Diff., Logs
12	First Diff., Logs	62	No Change	112	First Diff., Logs	162	First Diff., Logs
13	First Diff., Logs	63	No Change	113	First Diff., Logs	163	First Diff., Logs
14	First Diff., Logs	64	No Change	114	First Diff., Logs	164	First Diff., Logs
15	First Diff., Logs	65	No Change	115	First Diff., Logs	165	First Diff., Logs
16	First Diff., Logs	66	No Change	116	First Diff., Logs	166	First Diff., Logs
17	First Diff., Logs	67	No Change	117	First Diff., Logs	167	First Diff., Logs
18	First Diff., Logs	68	No Change	118	First Diff., Logs	168	First Diff., Logs
19	First Diff., Logs	69	No Change	119	First Diff., Logs	169	First Diff., Logs
20	First Diff., Logs	70	No Change	120	First Diff., Logs	170	First Diff., Logs
21	First Diff., Logs	71	No Change	121	First Diff., Logs	171	First Diff., Logs
22	First Diff., Logs	72	No Change	122	First Diff., Logs	172	First Diff.
23	First Diff., Logs	73	No Change	123	First Diff., Logs	173	First Diff.
24	First Diff., Logs	74	No Change	124	First Diff., Logs	174	First Diff.
25	First Diff., Logs	75	No Change	125	First Diff., Logs	175	First Diff.
26	First Diff., Logs	76	No Change	126	First Diff., Logs	176	First Diff.
27	First Diff., Logs	77	No Change	127	First Diff., Logs	177	First Diff., Logs
28	First Diff., Logs	78	No Change	128	First Diff., Logs	178	First Diff., Logs
29	First Diff., Logs	79	No Change	129	First Diff., Logs	179	First Diff.
30	First Diff., Logs	80	No Change	130	First Diff., Logs	180	First Diff.
31	First Diff., Logs	81	No Change	131	First Diff., Logs	181	First Diff.
32	First Diff., Logs	82	No Change	132	First Diff., Logs	182	First Diff.
33	First Diff., Logs	83	No Change	133	First Diff., Logs	183	First Diff.
34	First Diff., Logs	84	No Change	134	First Diff., Logs	184	First Diff.
35	First Diff., Logs	85	No Change	135	First Diff., Logs	185	First Diff.
36	First Diff., Logs	86	No Change	136	First Diff., Logs	186	First Diff.
37	First Diff., Logs	87	No Change	137	First Diff., Logs	187	First Diff.
38	First Diff., Logs	88	No Change	138	First Diff., Logs	188	First Diff.
39	First Diff., Logs	89	No Change	139	First Diff., Logs	189	First Diff., Logs
40	First Diff., Logs	90	No Change	140	First Diff., Logs	190	First Diff., Logs
41	First Diff., Logs	91	No Change	141	First Diff., Logs	191	First Diff., Logs
42	First Diff., Logs	92	No Change	142	First Diff., Logs	192	First Diff., Logs
43	First Diff., Logs	93	No Change	143	First Diff., Logs	193	First Diff., Logs
44	First Diff., Logs	94	No Change	144	First Diff., Logs	194	First Diff., Logs
45	First Diff., Logs	95	No Change	145	First Diff., Logs	195	First Diff., Logs
46	First Diff., Logs	96	First Diff.	146	First Diff., Logs		
47	First Diff., Logs	97	First Diff.	147	First Diff., Logs		
48	First Diff., Logs	98	First Diff.	148	First Diff., Logs		
49	First Diff., Logs	99	No Change	149	First Diff., Logs		
50	First Diff., Logs	100	No Change	150	First Diff., Logs		

Bibliography

Alvarez R., Camacho M., Perez-Quiros G. (2012), Finite sample performance of small versus large scale dynamic factor models, Centre for Economic Policy Research, Discussion Paper 8867.

Aruoba B., Diebold F., Scotti C. (2009), Real-time measurement of business conditions, Journal of Business and Economic Statistics, 27, 417-427.

Aruoba B., Diebold F. (2010), Real-time macroeconomic monitoring: Real activity, inflation and interactions, American Economic Review, 100, 20-24.

Bai J. (2003), Inferential theory for factor models of large dimension, Econometrics, 71, 135-173.

Bailey N., Kapetanios G., Pesaran H.M. (2012), Exponent of cross-sectional dependence: estimation and inference, CESifo Working Paper, No. 3722.

Ng S., Boivin J. (2006), Are more data always better for factor analysis?, Journal of Econometrics, 132, 169-194.

Brüggemann R., Krolzig H.M., Lutkepohl H. (2003), Comparison of Model Reduction Methods for VAR Processes, Technical Report 2003-W13, Nuffield College, University of Oxford.

Buchen T., Wohlrabe K. (2011), Forecasting with Many Predictors:Is Boosting a Viable Alternative?, Economics Letters, 113, 16-18.

Bulligan G., Marcellino M.,Venditti F. (2011), Forecasting economic activity with targeted predictors, Mimeo.

Camacho M., Perez-Quiros G. (2010), Introducing the Euro-STING: Short Term INdicator of Euro Area Growth, Journal of Applied Econometrics, 25, 663-694.

Camba-Mendez G., Kapetanios G., Smith R. J., Weale M. R. (2001), An automatic leading indicator of economic activity: forecasting GDP growth for European countries, Econometrics Journal, 4, S56-S90.

Caggiano G., Kapetanios G., Labhard V. (2011), Are more data always better for factor analysis? Results for the Euro Area, the six largest Euro Area countries and the UK, Journal of Forecasting, 30, 8, 736-752.

Carriero A., Kapetanios G., Marcellino M. (2011), Forecasting Large Datasets with Bayesian Reduced Rank Multivariate Models, Journal of Applied Econometrics, 26, 733-761,

De Mol C., Giannone D., Reichlin L. (2006), Forecasting with a Large Number of Predictors: Is Bayesian Regression a Valid Alternative to Principal Components, Discussion Paper 5829, Centre for Economic Policy Research.

Fernandez C., Ley E., Steel M.F.J. (2001), Benchmark Priors for Bayesian Model Averaging, Journal of Econometrics, 100, 381-427.

Forni M., Hallin M., Lippi M., Reichlin L. (2000), The Generalised Factor Model: Identification and Estimation, Review of Econmics and Statistics, 82, 540-554.

Forni M., Hallin M., Lippi M., Reichlin L. (2005), The Generalised Factor Model: One-Sided Estimation and Forecasting, Journal of the American Statistical Association, 100-471, 830-840.

Foroni C., Marcellino M. (2011), A Comparison of Mixed Frequency Approaches for Modelling Euro Area Macroeconomic Variables, Mimeo.

Foroni C., Marcellino M., Schumacher C. (2011), MIDAS and Unrestricted MIDAS models, Mimeo.

Goffe W.L., Ferrier G.D., Rogers J. (1994), Global Optimisation of Statistical Functions with Simulated Annealing, Journal of Econometrics, 60-1, 65-99.

Hajek B. (1998), Cooling Schedules for Optimal Annealing, Mathematics of Operations Research, 13-2, 311–331.

Hartl H.R.F., Belew R.K. (1990), A Global Convergence Proof for a Class of Genetic Algorithms, Technical Report, Technical University of Vienna.

Helland I.S. (1988), On the Structure of Partial Least Squares Regression, Communications in Statistics - Simulation and Computation, 17, 581-607.

Helland I.S. (1990), Partial Least Squares Regression and Statistical Models, Journal of Statistics, 17, 91-114.

Hendry D.F. (1995), Dynamic Econometrics, Oxford University Press.

Hendry D.F. (1997), On Congruent Econometric Relations: A Comment, Carnegie-Rochester Conference Series on Public Policy, 47, 163–190.

Hoover K.D., Perez S.J. (1999), Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Set, Econometrics Journal, 2, 167-191.

Hannan E.J., Quinn B.G. (1979), The Determination of the Order of an Autoregression, Journal of the Royal Statistical Society (Series B), 41, 190-195.

Kapetanios G. (2006), Variable Selection in Regression Models using Non-Stantard Optimisation of Information Criteria, Computational Statistics and Data Analysis, 52-1, 4-15.

Kapetanios G., Marcellino M., Papailias F. (2012), Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods, Unpublished Working Paper.

Kapetanios G., Marcellino M., Papailias F. (2012), Forecasting EU economic activity using financial conditions indexes, Unpublished Working Paper.

Kapetanios G., Marcellino M., Papailias F. (2012), Business cycles dating for EU economies: an empirical search for the optimal settings, Unpublished Working Paper.

Kapetanios G., Marcellino M., Papailias F. (2012), Forecasting EU economic activity using summary indicators, Unpublished Working Paper.

Krozlig H.M., Hendry D.F. (2001), Computer Automation of General-to-Specific Model Selection Procedures, Journal of Economic Dynamics and Control, 25 (6–7), 831-886.

Kuzin V., Marcellino M., Schumacher C. (2011), MIDAS vs. Mixed-Frequance VAR: Nowcasting GDP in the Euro Area, International Journal of Forecasting, 27, 529-542.

Marcellino M., Schumacher C. (2010), Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP, Oxford Bulletin of Economics and Statistics, 72, 518–550.

Morinaka Y., Yoshikawa M., Amagasa T. (2001), The L-Index: An Indexing Structure for Efficient Subsequence Matching in Time Sequence Databases, Proceedings of the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining.

Pesaran H.M. (2012), Testing weak cross-sectional dependence in large panels, CESifo Working Paper, No. 3800.

Schwarz G. (1978), Estimating the Dimension of the Model, Annals of Statistics, 6, 461-464.

Sin C.Y., White H. (1996), Information Criteria for Selecting Possibly Misspecifed Parametric Models,Journal of Econometrics, 71-(1-2), 207-225.

Stock J., Watson M. (1989), New indexes of coincident and leading economic indicators. *NBER Macroeconomic Annual*, 351-394.

Stock J., Watson M. (1991), A probability model of the coincident economic indicators. In Leading Economic Indicators: New Approaches and Forecasting Records, edited by K. Lahiri and G. Moore. Cambridge University Press.

Stock J., Watson M. (1998), Diffusion Indexes, NBER Working Paper, No. 6702 .

Stock J., Watson M. (1999), Forecasting Inflation, Journal of Monetary Economics, 44, 293-335.

Stock J., Watson M. (2002), Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association, 297, 1167-1179.

Stock J., Watson M. (2002), Macroeconomic Forecasting using Diffusion Indexes, Journal of Business & Economics Statistics, 20, 147-162.

Stock J., Watson M. (2006), Forecasting with Many Predictors, In: Graham, E., Granger, C., Timmerman, A. (Eds.), Handbook of Economic Forecasting, Amsterdam, 546-550;

Wold H. (1982), Soft Modeling. The Basic Design and Some Extensions, In: Joreskog, K. G. and H., Wold, Systems Under Indirect Observation, Vol. 2, North-Holland, Amsterdam ;

III
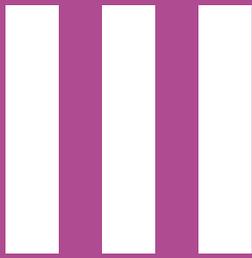
Composite Indicators to Anticipate Cyclical Movements

7

An Overview of Cyclical Indicators

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ludovic Calès
European Commission, Joint Research Centre (JRC)
ludovic.cales@ec.europa.eu

Table of Contents

7.1 Introduction

This part describes the most commonly used classes of composite indicators which aim to mimic in real-time, or to anticipate, the evolution of an unobserved component representing the cyclical movements of a given macroeconomic time series. Those indicators are directly stemming from the seminal work of Burns and Mitchell (1946) and of the subsequent activities carried out by the National Bureau of Economic Research (NBER). They can also be viewed as a substantial improvement and modernisation of the synthetic economic barometers developed right before the Second World War. In this context, the indicators presented in this part provide an estimation of a continuous variable which represents the cyclical movements based on a given definition of cycle. Obviously, those indicators can also produce, as a by-product, an estimation of the sequence of turning points under the condition that an appropriate dating rule is applied to the estimated cyclical components (see Anas et al. (2017) chapter 14, section 14.2, as well as Harding (1997), Harding and Pagan (2002), Harding and Pagan (2011) and Chauvet and Hamilton (2006) for more details). The main issues to be considered when constructing cyclical composite indicators can be synthesised as follows: the identification of the reference variable, the choice of the reference cycle, the selection of the statistical method used for selecting variables and for estimating and forecasting the cyclical component.

The chapter is structured as follows: Section 7.2 discusses the problems related to the identification of the reference variable while section 7.3 deals with the choice of the reference cycle; section 7.4 discusses statistical methods related to the matter, section 7.5 provides a literature review on the subject. Section 7.6 describes existing leading indicators, section 7.7 provides a guide to the following chapters and section 7.8 concludes.

7.2 Identification of the reference variable

This is a crucial step in the construction of cyclical indicators because the cyclical components of the statistical variable can substantially differ in terms of volatility, amplitude and phase. The most appropriate reference variable should provide an accurate and not too volatile description of the cyclical movements of the economic system as a whole. For this reason, in the NBER based approach, the reference variable is not an individual macroeconomic variable but a latent one obtained by averaging a number of relevant macroeconomic indicators.

Alternatively, individual macroeconomic indicators, such as the industrial production index, have been widely used as a reference variable. Nowadays, in presence of a generalisation of economic fluctuations, which characterise almost all sectors of economic activity, the role of industrial production has progressively declined and the GDP seems to be the most appropriate candidate. This consideration is especially appropriate for developed economies, while for emerging and developing economies it is likely to assume that the GDP can still play a central role as reference variable. By contrast, for underdeveloped economies, the situation can be much more complex. Even if we do not have any direct experience in compiling composite indicators for those economies, we can assume that the reference variable can be represented by an indicator reflecting specific characteristics of the considered economy. Such indicators could be either related to mining and quaring activities, production and export of raw materials or energy products or even agricultural production. Coming back to the situation where the GDP is the ideal reference variable, we have to notice that since the GDP, is in most cases, only available at quarterly frequency, this constitutes an important limitation to its use. Monthly proxies of GDP could be used instead but since they are not necessarily publicly available, their utilisation as a reference variable could impair the transparency of the indicators and their replicability.

As an example, in the composite leading indicator (CLI) produced by the OECD, (see Astolfi et al., chapters 10 and 19), the reference variable for the quasi totality of countries is represented by a monthly interpolated version of the GDP with the exception of China, where the industrial production remains the reference variable. Finally, other cases where the reference variable can differ from the GDP or the IPI are represented by sectoral

indicators. Those indicators are designed to study the cyclical movements of the specific economic sector (e.g. labour market) or branch of activity. In this case, the reference variable can be represented by an employment indicator or by the output or the added value of a given industry. An example of such indicator is constituted by "EU Craft and SME Barometer" published by the UEAPME (the employers' organisation representing the interests of European crafts, trades and SMEs at EU level).

7.3 Choice of the reference cycle

The indicators developed by the NBER, and now regularly produced by the Conference Board (see Ozyildirim, Chapter 8), use the Burns and Mitchell definition of business cycle which statistically corresponds to the trend-cycle component, obtained by filtering out the irregular component from the seasonal adjusted version of a macroeconomic variable, see Ozyildirim (2017), Chapter 8. Alternatively, the OECD has developed a set of indicators using the output-gap as a reference cycle, see Gyomai et al. (2017), Chapter 10. This means that, after the seasonal adjustment, an additional filter has to be used to separate the trend from the cyclical and, possibly, the irregular components of seasonal adjusted series.

In this case, end-point estimates problems due to the application of the de-trending filters might appear. They constitute the most relevant drawback of this category of indicators (see Mazzi et al. (2017), Chapter 8). On the other hand, the choice of the reference cycle proposed by the NBER is often not very informative due to the long expansionary phase having characterised, especially in the past, the economic behaviour.[1] It is also important to underline that cyclical composite indicators, having the output-gap as the reference cycle, are particularly important also for central banks both for business cycle analysis and for inflation monitoring.

When using a composite indicator having the output gap as the reference cycle we obtain clearer information on the various phases of economic growth (i.e. above or below the trend) which are only indirectly provided by indicators based on the classical business cycle definition. On the other hand, composite indicators based on the output gap do not provide any information on the occurrence of recessions and on the exit from them. Ideally, the availability of both types of indicators described in this section could supply a better picture of the cyclical movements provided that a continuous monitoring of them is carried out in order to avoid possible inconsistencies.

7.4 Selection of the statistical methods

Once the reference variable and cycle have been identified, the composite indicators can be built up in two main ways: by pooling carefully selected variables or by deriving it from a model.

The first approach consists in pre-selecting variables showing cyclical patterns conform to the reference one. The selected variables can have either a coincident or a leading behaviour in comparison to the reference cycle, allowing for estimating in real-time or anticipating cyclical movements. These variables are then aggregated into composite indicators to infer the cyclical components of the reference variable. The selection of the variables can also use the information provided by historical turning points chronologies, whenever available. In particular the selection should focus on their consistent timing (lagging/coincident/leading), conformity, smoothness and economic significance. Variable selection can be carried out either by using non-parametric or parametric techniques (see Guégan (2017), chapter 5).

As illustrations, the sets of macroeconomic variables considered by the Conference Board (formerly the U.S. Department of Commerce) and the OECD are presented in Chapter 8, and in chapter 10, respectively. In order to assess the conformity of these variables, they are either decomposed into their trend and cyclical

---

[1] Even though the Conference Board also considers two sub-periods: 1959-1983 and 1984-present, see Chapter 8

components (if the reference cycle is the output gap) or they are filtered to remove the irregular component (if the reference cycle is the classical business cycle proposed by Burns and Mitchell). However, their trends and cycles are not directly observable and even if economic theory can provide some clues on the "true" processes driving the trend and cycle, there is no consensus on what should be their correct estimates. It leads to a wide variety of models and techniques to estimate them. The de-trending methods can be parametric or non-parametric, uni- or multivariate, linear or non-linear (see Mazzi et al. (2017), chapter 9).

Finally, the selected variables are aggregated into composite indicators. The composite indicator is then a weighted sum/average of components which are usually standardized to deal with the different units in which they are measured. Variables can also be subject to further adjustment to remove the excessive volatility. The weighting scheme is built up in order that the weights attached to each variable component sum to one. Weighting schemes can considerably differ and they can lead to very different results. They can vary from purely subjective ones to those strongly based on statistical considerations (see Carriero et al. (2017), chapter 3).

The second approach consists in extracting, from a given model, a latent variable which is interpreted as a leading or coincident proxy of the reference variable and cycle. The most frequently used modelling techniques for constructing cyclical composite indicators can be grouped in the following categories:

- the linear regression models,

- the VAR and VECM models,

- the unobserved components models,

- the dynamic factor models,

- the Structural VAR decomposition,

- the DSGE models.

It is important to observe that the last two methods, and DSGE in particular, require strong economic assumptions implying an a priori decision on what kind of macroeconomic approach we want to privilege. On the other hand, the first four methods are much more based on statistical evidence and can be assimilated to filtering techniques aiming to extract a cyclical signal. Finally, by selecting certain models such as the dynamic factor models or the unobserved components ones, the variable selection and the estimation steps can be merged. By contrast, other methods still require a preselection of the variables, which can also be carried out within the chosen modelling techniques (VAR, linear regression), i.e. by using a general-to-specific (GETS) modelling, see Guégan (2017), chapter 5.

7.5 Review of the literature

The seminal work of Mitchell and Burns (1938) and Burns and Mitchell (1946) began by locating turning points in many series measuring, partially, the aggregate economic activity and thus defining specific cycles. In a second step, they aggregate this information to define a single set of turning points which they defined as the reference cycle. As such, they defined an index of conformity to compare the cycles of individual series to the reference cycle. The procedure followed by Burns and Mitchell was considered by Harding and Pagan (2002) as being "more an art than a science". And the pooling approach used by practitioners such as the Conference board (see Hertzberg and Beckman (1989), Green and Beckman (1993) and The Conference Board (2001) and the OECD (see OECD (1987) and Gyomai and Guidetti (2012)) stems directly from their work.

This approach has been refined by Bry and Boschan (1971) which proposed a set of "censoring rules" making the turning points satisfy criteria of duration, length of phases and complete cycles. Along the years, they made more use of de-trending methods such as the moving average filter, the Hodrick-Prescott and

Christiano-Fitzferald filters, see Hodrick and Prescott (1980) and Christiano and Fitzgerald (2003), and the phase averaging trend, see Boschan and Ebanks (1978). Finally, Harding and Pagan (2002) proposed a statistical foundation for Burns and Mitchell's approach, defining clearly the properties of the cycles (duration of the cycle and its phases, amplitude of the cycle and its phases, any asymmetric behavior of the phases, cumulative movements within phases).

In parallel to the practitioners, academics focused much more on the study of the co-movements among the series instead of focusing on the construction of the reference cycle, see Koopmans (1947). And quantitative measures extracted from parametric models where proposed. Harvey (1985) introduced the unobserved components models, Blanchard and Quah (1989) proposed the Structural VAR, Stock and Watson (1989), Stock and Watson (1991) and Stock and Watson (1992) presented the dynamic factor models. The DSGE models introduced by Woodford (2003) and Smets and Wouters (2003) are frequently used by central banks.

The $\alpha AB\beta CD$ approach first presented in Anas and Ferrara (2004) and further developed in Anas et al. (2007) and Anas et al. (2008) proposed a clear decomposition of the cycles, identifying the turning points by means of latent variables representing the different phases. It aims to reconcile the two approaches presented here and is introduced in Anas et al. (2017), chapter 14.

7.6 On existing leading indicators

Providing an overview of existing cyclical coincident or leading indicators is almost impossible due to the high number of them developed all around the world both by private and public institutions as well as by academia. In this section we are briefly mentioning some of the most known indicators without pretending to be exhaustive nor to cover all geographic areas.

The wider range of cyclical indicators for the business cycle and the growth cycle, is proposed by the Conference Board and the OECD respectively. The main advantage of those indicators is that they allow for a sound cross-country comparison since they are based on the same methodology even if the components can differ according to countries' specificities. Also the Economic Cycle Research Institute (ECRI) is providing a wide variety of composite, coincident and leading indicators for a large number of countries and economic areas. They provide leading indicators for inflation and long leading indicators for the cyclical situation. At the euro area level we would like to mention the euro Area-wide Leading Indicator (ALI) produced by the ECB (see De Bondt and Hahn (2014)) but also the euro-coin indicator produced by Bank of Italy (see Altissimo et al. (2007)) and the DZ-Euroland indicator. When looking at national indicators, it is almost impossible to make a list for the most relevant coincident and leading indicators. Nevertheless, at the euro area level we would like to mention the indicators produced by the INSEE France, ISTAT Italy, as well as by the Netherlands which uses the business cycle tracer to construct a composite coincident indicator of the growth cycle. Finally, outside Europe we would like to mention the composite coincident and leading indicators developed by INEGI Mexico and by the Statistical Office of South Korea mainly because they show the importance of the direct involvement of statistical authorities on the compilation of such kinds of indicators.

7.7 Guide to the part

As seen in Section 7.3, indicators can be constructed by using a simple pooling technique or in a model based way. If the user is interested in the second approach, he can refer to Moauro (2017) (chapter 11), where the dynamic factor models introduced by Stock and Watson (1989)are presented along with their improvements and extensions. If the user privileges the Burns and Mitchell approach, he can refer to Ozyildirim (2017) (chapter 8), while Gyomai et al. (2017) (chapter 10) describes the output gap based version of this approach. Mazzi et al. (2017) (chapter 9) reviews the vast literature on de-trending methods, analysing a large number

of them and presenting their respective advantages and drawbacks. Even if de-trending techniques can be of some interest also for the Conference Board approach (mainly for the possibility of using some of them as filters to remove noise), they are surely essential for constructing cyclical composite indicators using the output gap as the reference cycle. In fact, the output gap is obtained by removing the trend and, possibly, the irregular component of the GDP or any other indicator measuring the economic production. This decomposition is achieved by using a number of filtering techniques such as those described in chapter 9. For this reason, the chapter on de-trending techniques is preparatory for the one on the OECD method. The rest of this section presents a short summary of the main contents of the chapters included in this part.

– Chapter 8 presents the production process of composite indicators at the Conference Board. Following Burns and Mitchell (1946), a wide set of indicators are considered to capture the economic activity. They are classified according to their relationship to a reference chronology, particularly focusing on their consistent timing (lagging/coincident/leading), conformity, smoothness, economic significance and currency. Then they are combined into composite indicators. Along the years different aggregation methods have been experimented and their experience is presented.

– Chapter 9 reviews the vast literature on de-trending methods. They are presented according to their characteristics: linear/non-linear, parametric/non-parametric, structural/non-structural and their statistical fit. All the methods introduces in Section 7.3 are reviewed and an application to the de-trending of the euro-area GDP is presented.

– Chapter 10 presents the production process of composite indicators at the OECD. As with the Conference Board, the pooling method is preferred. Their Composite Leading Indicators (CLIs) focus on the correct identification of the phases and, as such, they are not optimized to provide accurate forecasts. Their experience is presented.

– Chapter 11 presents the dynamic factor models introduced by Stock and Watson (1989), along with its improvements and extensions. A focus is given on two of its extensions: the EUROMIND model developed in Frale et al. (2011) which leans toward mixed frequency models, and the EUROMIND-S model developed in Frale et al. (2010) which leans toward multi-factors models. Both models are currently experimented by Eurostat.

7.8 Conclusions

After having highlighted the problems related to the identification of the reference variable, this chapter dealt with the choice of the reference cycle. Its aim was to guide the reader through the different classical approaches in constructing cyclical composite indicators, introducing the approaches presented in the following chapters of this part.

Bibliography

Altissimo F., Cristadoro R., Forni M., Lippi M. and Veronese G. (2007), New Eurocoin: Tracking economic growth in real time, The Review of Economics and Statistics, No. 92, 1024-1034.

Anas J., Billio M., Ferrara L. and Mazzi G.L. (2017), Chapter 14: Composite cyclical indicators detecting turning points within the ABCD framework, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Anas J. and Ferrara, L. (2004), Turning points detection: The ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, 1, 1-36.

Anas J., Billio M., Ferrara L. and LoDuca M. (2007), Business cycle analysis with multivariate Markov Switching models, in Growth and Cycle in the Euro zone, G.L. Mazzi and G. Savio (eds.), 249-260, Palgrave MacMillan, New York.

Anas J., Billio M., Ferrara L. and Mazzi G.L. (2008), A system for dating and detecting cycles in the Euro area, The Manchester School, 76, 549-577.

Blanchard O.J. and Quah D. (1989), The dynamic effects of aggregate demand and supply disturbances, American Economic Review, 79, 655-73.

Boschan C. and Ebanks W.W. (1978), The Phase-Average Trend: A New Way of Measuring Growth, in 1978 Proceedings of the Business and Economic Statistics Section, American Statistical Association, Washington, D.C.

Bry G. and Boschan C., (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, New York, NBER.

Burns, A.F. and Mitchell, W.C. (1946), Measuring Business Cycles, National Bureau of Economic Research, New York, 1946.

Carriero A, Marcellino M. and Mazzi G.L. (2017), Chapter 3: Definitions and Taxonomy of indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Chauvet M. and Hamilton J.D. (2006), Dating business cycle turning points, Contributions to Economic Analysis 276, 1-54.

Christiano L. and Fitzgerald T.J. (2003), The Band Pass Filter, International Economic Review, 44, 2, 435-465.

De Bondt G.J. and Hahn E. (2014), Introducing the Euro Area-wide Leading Indicator (ALI): Real-Time Signals of Turning Points in the Growth Cycle from 2007 to 2011. Journal of Forecasting, 33, 47âĂŞ68.

The Conference Board (2001), Business Cycle Indicators Handbook, New York, NY.

Frale C., Marcellino M., Mazzi GL. and Proietti T. (2010), Survey Data as Coincident or Leading Indicators, Journal of Forecasting, 29, 109-131.

Frale C., Marcellino M., Mazzi G.L. and Proietti T. (2011), EUROMIND: a monthly indicator of the euro area economic conditions, Journal of the Royal Statistical Society, series A, 174, part 2, 439-470.

Green G.R., Beckman M.A. (1993), Business cycle indicators: upcoming revisions of composite leading indicators, Survey of Current Business, October, Vol. 73, pp. 44-51.

Guégan D. (2017), Chapter 5: Review of Parametric and non-Parametric Variable and Model Selection Techniques, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Gyomai G. and Guidetti E. (2012), OECD System of Composite Leading Indicators. OECD, Statistics Directorate. Paris: OECD.

Gyomai G., Ahmad N. and Astolfi R. (2017), Chapter 10: The OECD System of Composite Leading Indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Harding D. (1997), The definition, dating and duration of cycles, paper presented to the Melbourne Institute conference on Business cycle policy and analysis.

Harding D. and Pagan A. (2002), Dissecting the cycle: a methodological investigation. Journal of Monetary Economics, 49(2), pp. 365-381.

Harding D. and Pagan A.R. (2011), Econometric Analysis and Prediction of Recurrent Events, No 75, NCER Working Paper Series from National Centre for Econometric Research.

Harvey A.C. (1985), Trends and Cycles in Macroeconomic Time Series, Journal of Business and Economic Statistics, vol. 3, pp. 216-230.

Hertzberg, M.P. and Beckman B. A. (1989), Business Cycle Indicators: Revised Composite Indexes, Survey of Current Business 72 (6), pp. 23-28.

Hodrick R.J. and Prescott E.C. (1980), Postwar U.S. Business Cycles: An Empirical Investigation, Carnegie Melon University discussion paper no. 451.

Koopmans T.J., (1947), Measurement without Theory, Review of Economics and Statistics, 29, 161-172.

Mazzi G.L., Mitchell J. and Ozyildirim A (2017), Chapter 9: Alternative de-trending methods, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mitchell W.C. and Burns A.F. (1938), Statistical Indicators of Cyclical Revivals, National Bureau of Economic Research, Bulletin 69.

Moauro F. (2017), Chapter 11: The Stock and Watson approach to composite indexes of coincident and leading indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

OECD (1987), OECD Leading Indicators and business-cycles in Member Countries 1960-1985. OECD, Department of Economics and Statistics. Paris: OECD.

Ozyildirim A. (2017), Chapter 8: Business Cycle Indicator Approach at the Conference Board, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Smets F. and Wouters R. (2003), An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area". Journal of the European Economic Association, 1 (5): 1123-1175.

Stock J.H. and Watson M.W. (1989), New indexes of coincident and leading indicators, in Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA, pp. 351-394.

Stock J.H. and Watson M.W. (1991), A probability model of the coincident economic indicators, in Lahiri, K., Moore,G.H. (Eds.), Leading Economic Indicators: New Approaches and Forecasting Records. Cambridge University Press, Cambridge, UK, pp. 63-89.

Stock J.H. and Watson M.W. (1992), A procedure for predicting recessions with leading indicators: econometric issues and recent experience, NBER Working Papers Series No. 4014.

Woodford M. (2003), Interest and prices: Foundations of a theory of monetary policy. Princeton University Press, Princeton, NJ.

8

Business Cycle Indicator Approach at The Conference Board

Author

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

Table of Contents

8.1 Introduction

Business cycle indicators and related composite indexes have proven to be useful tools for analyzing alternating sequences of economic expansion and contraction known as business cycles. The indicator approach to business cycle analysis and forecasting originated in the 1930s at National Bureau of Economic Research (NBER) with the work of Wesley Mitchell and Arthur Burns (see the list of references for relevant works).

Over subsequent decades the approach was developed and refined, mostly at the NBER under the leadership of Geoffrey H. Moore. Starting in the late 1960s, the U.S. Department of Commerce Bureau of Economic Analysis (BEA) began publishing the business cycle indicator data and composite indexes of leading, coincident, and lagging indicators. In late 1995, the Business Cycle Indicators program was privatized, and starting in 1996, The Conference Board took over the responsibility of maintaining the database and publishing the monthly report.

This chapter presents an overview of the indicator approach as developed at the NBER and now continued at The Conference Board. The main focus of the chapter is on the selection and evaluation of economic indicators that are used as components of composite business cycle indexes which aid the measurement and analysis of the "classical" business cycle. This indicators approach has also been extended subsequently to utilize such composite indexes in the analysis of growth cycles and growth rate cycles (Klein and Moore , 1985; Mintz , 1969, 1972, for example).

8.2 How are Indicators Selected

Over the years, many economic indicators have been evaluated and classified based on a set of criteria which focuses on their properties vis-à-vis the business cycle. In order to determine the cyclical properties of a series, it is compared to an established reference chronology which specifies the peaks and troughs of the business cycle. (In the United States, the NBER committee on dating business cycles determines and maintains this chronology).

– The first criterion is **Consistent Timing** — the series must exhibit a consistent timing pattern as a leading, coincident or lagging indicator. A measure of timing consistency is the standard deviation of the leads or lags of the turning points of the series relative to the business cycle turning points.

– Another criterion is **Conformity** — the series must conform well to the business cycle. A series that conforms to the business cycles perfectly would not miss any of the business cycle turning points and would not exhibit any extra cycles. A reference turning point is said to be missed by the series if the series does not have a closely related turning point in the "near" vicinity of the reference turning point (the appropriate closeness of the turning point in the series to the reference cycle depends on whether the series can be classified as leading, long leading, or lagging). Determination of misses relies crucially on the ability to match a peak (trough) in the series with a peak (trough) in the business cycle without an intervening trough (peak). An extra cycle occurs when the series has a peak and a trough not associated with the reference chronology.

– A third criterion is **Smoothness** — month-to-month movements must not be too erratic. A good measure of smoothness is the standard deviation of the monthly log differences of the series. Other measures include months for cyclical dominance (MCD) and coefficient of variability of the monthly changes.

– Fourth, it is important that the series have **Economic Significance** — cyclical timing must be economically meaningful and logical.

– Fifth, the series must have **Statistical Adequacy** — data must be collected and processed in a statistically reliable way.

– Finally, a sixth criterion is **Currency** or **Timeliness** — series must be published on a reasonably prompt schedule, preferably within a month.

When these standards are strictly applied, relatively few individual time series pass muster. No quarterly series qualifies for lack of currency and many monthly series lack smoothness. Indeed, there is no single time series that fully qualifies as an ideal cyclical indicator. This is one motivation for combining various indicators into composite indexes as described below in the subsection on composite indexes.

8.2.1 Classification by Cyclical Timing

Starting with the work of Arthur Burns and Wesley Mitchell, the NBER program on cyclical fluctuations developed a set of procedures to define and measure cycles in the business cycle frequency along with the concept of business cycles. For example, phases (contraction or expansion) whose duration is less than 5 months are not considered as belonging to the business cycle frequency. Similarly, cycles (contraction plus expansion) whose duration is less than 15 months are not considered business cycles. Bry and Boschan (1971) developed a computer algorithm for selecting turning points which formalized and automated this set of definitions and procedures. While the algorithm is extremely useful in the evaluation, business cycle indicators and indexes, it should be emphasized that its results should always be reviewed as special cases, outliers, and/or unusual circumstances, could interfere with the correct identification of turning points.

Based on these definitions, once their turning points are determined, cyclical indicators are classified into three categories – leading, coincident, and lagging – based on the timing of their movements. Coincident indicators, such as employment, production, personal income, and real manufacturing, and trade sales, are broad series that measure aggregate economic activity; thus, they define the business cycle. This is why the NBER Business Cycle Dating Committee relies heavily on these series to determine when recessions begin and end (In the international portfolio of countries covered by The Conference Board, the reference cycle is determined by looking at the turning points of a composite index constructed from a combination of their coincident index components and real GDP. In a few cases, such as Brazil's CODACE, independent committees determine and announce recession dates.).

Leading indicators, such as average weekly hours, new orders, consumer expectations, housing permits, stock prices, and interest rate spread, are those series that tend to shift direction in advance of changes in economic activity or the reference cycle. Because these series contain predictive information, they get the lion's share of the attention. However, understanding the current state of the economy through the analysis of the coincident indicators is equally valuable; indeed, this should be the first step in any forecasting exercise. In fact, an index of current economic conditions such as the Coincident Economic Index (CEI) can be thought of as the cornerstone of the indicator approach because it is instrumental in measuring business cycles and defining the reference chronology. Therefore, the development of a system of business cycle indicators almost always begins with an analysis of coincident indicators.

It is important to recognize that leading indicators are more meaningful when used within a framework of a system of cyclical indicators – including coincident and lagging indicators that define and describe business cycles. Even lagging indicators which receive much less attention have a role in confirming that a cyclical phase is over and that the next phase has begun. Moreover, another less cited index, the ratio of the coincident index to lagging index is often thought of as a leading indicator since it measures the level of current economic activity relative to the lagging index, which is thought of as a summary measure of the cost of doing business.

8.2.2 Classification by Economic Process

The U.S. Department of Commerce published subgroup indexes of leading indicators for 5 subgroups, all of which are represented in the leading index. These subgroups are

1. marginal employment adjustments: changes in the average work week, claims for unemployment insurance, plans to hire, help-wanted advertising, etc.

2. capital investment commitments: new orders for nondefense capital equipment, orders for plant and equipment, housing permits, etc.

3. inventory investment and purchasing: new orders for consumer goods and materials, vendor performance, change in manufacturing and trade inventories, change in sensitive materials prices, etc.

4. profitability: stock prices, corporate profits after tax, ratio of implicit price deflator to unit labor cost, etc.

5. money and financial flows: real money supply (M2), change in business and consumer credit outstanding, change in total liquid assets, bond yields, etc.

The indicators used in these subgroups come from a larger data set of more than 270 monthly and quarterly cyclical indicators maintained as the US Business Cycle Indicators (BCI) database. The components of the composite indexes of business cycle indicators were selected from this larger data set. An important function of the composite indexes is to highlight one of the most important aspects of business cycles, namely, the comovement of economic indicators across diverse aspects of economic activity at the macro level. Thus, one of the important elements of selection of the index components is to cover the diverse aspects of economic activity represented in these 5 subgroups. It is important to keep in mind that although the indexes summarize and emphasize the comovements that define business cycles, they should not be interpreted independently of their individual components, or, indeed, independently of other cyclical indicators.

8.3 Composite Indexes

In order to emphasize the cyclical patterns in the data and de-emphasize the volatility of individual indicators, the best of them are combined into composite indexes — specifically, into three separate indexes made up of leading, coincident, and lagging indicators. As mentioned above, one motivation for combining several indicators into one composite index is the observation that there is no single ""best" indicator. Another important motivation is to have a measure of diffusion among several indicators since business cycles are phenomena that are observed nearly simultaneously across all aspects of the economy.

These composite indexes serve as summary measures of the current and historical behavior of the cyclical indicators. Because they are averages, they tend to smooth out some of the volatility of individual series.

Use of composite indexes is consistent with the traditional view of the business cycle developed by Burns and Mitchell. In particular, composite indexes can reveal common turning point patterns in a set of economic data in a clearer and more convincing manner than the behavior of any individual component.

While the composite indexes and their components get the most attention, there is a wide range of closely related series maintained in the BCI database. Some of these series are substitutes, some are alternative, and some are complements. These series ought to be monitored to get a better sense of the plausibility or dependability of the information contained in the main indicators. In other words, if all series that are substitutes or closely related point in the same direction, then one can put more confidence in the signal given by that series. On the other hand, in some circumstances, divergences in such series could point to important developments in the economy which would not be otherwise apparent.

8.4 Aggregating the Cyclical Indicators into Composite Indexes

The idea of a composite index is an extension of diffusion indexes which measure the proportion of components that are rising. In contrast to diffusion indexes, the composite indexes also measure the magnitude of the change in the components. While diffusion indexes measure how widespread a cyclical movement is, composite indexes help measure the depth of that movement. Combined with the duration of the cyclical movement, these two types of indexes help determine whether observed changes in the data correspond to a persistent cyclical movement. The basic methodology for combining individual indicators into composites has remained largely the same since their inception. For details on various methodology changes see Hertzberg and Beckman (1989), Green and Beckman (1993), and The Conference Board (2001).

The monthly change in the index is an average of the monthly changes of its components.[1] Before aggregation, the monthly changes of the components are first adjusted by multiplying them with their standardization factor. This step is called a volatility adjustment. Standardization factors determine how monthly changes in each component contribute to the monthly change in the associated index. These factors are designed to give each component a similar opportunity to contribute to the change in the index in any given month. This adjustment equalizes the volatility of the contributions from each component in an index. The standardization factors are based on the inverse of the standard deviation of the monthly changes in the series.[2] These component standardization factors are made to sum to one. This summing to one of the standardization factors is done to assure that the cyclical part of the composite index is limited to a magnitude similar to the average deviation from the mean growth rate of the components of the index.

Once the volatility-adjusted monthly changes in the components are calculated, they are summed assuming equal weights for each component. Following the US Department of Commerce Bureau of Economic Analysis procedures, The Conference Board does not assign weights to individual components. This sum of the contributions is then cumulated, and after a trend adjustment to equate the long terms trends of the coincident economic index and the leading economic index, the resulting index is rebased to equal 100 in the base year (e.g. $2010 = 100$, the trend adjustment is only performed on the leading and lagging indexes).

8.4.1 Changes to Index Methodology

The original index aggregation methodology used a weighting scheme based on scores of the individual components derived from the six criteria for evaluating cyclical indicators. In a 1989 revision of the composite indexes this weighting scheme was dropped largely because the resulting weights were all very similar to each other. All of the weights were close to 1 and an unweighted index was not significantly different from a weighted one.

Also in 1989, the reverse-trend adjustment to equate the trends of the leading and lagging indexes to that of the coincident index was dropped, and instead a reverse-trend adjustment was used to equate the indexes' trends to that of GNP. In a 1993 revision, this reverse-trend adjustment was dropped.

Starting with the 1993 revision the standardization factors were normalized to sum to one. In 2001, the index standardization which equated the volatility of the leading and lagging indexes to that of the coincident index was dropped.

---

[1] The monthly change is calculated using a symmetric percent change formula. However, using percent changes or log differences makes little difference to the resulting index. If the component X is in percent change form or an interest rate, simple arithmetic differences are calculated: $x_t = X_t - X_{t-1}$. If the component is not in percent change form, a symmetric alternative to the conventional percent change formula is used: $x_t = 200 * (X_t - X_{t-1})/(X_t + X_{t-1})$. If the component X is a diffusion index (e.g. ISM New Orders Index) or an interest rate spread the monthly level is used as the monthly change in the component, $x_t = X_t$ (Diffusion indexes are first normalized by subtracting their sample mean and dividing by their standard deviation).

[2] See footnote 1.

8.4.2 Trend Adjustment of the Composite Indexes

As mentioned above, a (reverse) trend adjustment was made to the indexes earlier but dropped in the 1993 revision by the BEA.[3] The Conference Board continued to publish the composite indexes without a trend adjustment since it took over the responsibility for the indexes in 1995.

The trend of the leading economic index (LEI) (or the lagging index, LAG) is an artifact of the underlying series used to construct the composite index and lacks meaning on its own. Lack of an economically meaningful trend makes the LEI difficult to interpret and leads to large changes in the historic pattern of the LEI whenever changes in the components, which include both stationary and non-stationary series, are made. The reverse trend adjustment avoids these unnecessary changes in the historical record of the LEI and anchors the trend of the LEI to that of aggregate economic activity as measured by the coincident economic index or CEI.

Initially, the trend of CEI was used for the purpose of the trend adjustment by the U.S. Department of Commerce. The trend of the CEI appears to be the natural candidate for the measure of trend of economic activity to add to the LEI, because CEI is comprehensive, it is part of the indicators system, and it is monthly. By reverse trend adjustment using the trend of CEI, the LEI will adopt the meaningful trend of economic activity, and the trend adjusted LEI will facilitate comparisons between alternative versions of the LEI when components are changed.

The following is a simple method for accomplishing this adjustment:

- Divide each component by its standard deviation following current procedures to equalize the volatility of the individual components.[4]

- Sum these standardized (or normalized) monthly changes to calculate the monthly change in the composite index of the cyclical component of the LEI.

- Demean the monthly change of the LEI obtained in the previous step.

- Add the mean of the monthly percent changes in the CEI as the estimated trend component from CEI.

- Cumulate the LEI from this series and rebase it (e.g. 2010 = 100).

This is essentially what the current methodology does.

The method currently uses the mean of the monthly changes in CEI over the entire sample period to add as a trend to the LEI. Other ways to do this include similar linear trends but over sub-periods or nonlinear trends. For example:

- Calculating the mean of the monthly changes over two sub-periods (1959-1983 and 1984-present),

- Using a nonlinear trend such as the approximation to the Phase Average Trend using a Hodrick and Prescott (1997) trend (see also Ravn and Uhlig (2001)). This procedure may not be always feasible because of severe end-point problems, which would potentially give rise to large revisions in the estimation of the trend with the most recent data.

The most suitable approach to the trend adjustment of the composite index appears to be the full sample mean adjustment or a two sub-period (1959-1983 and 1984-present) procedure to calculate the trend although both of these approaches might have some drawbacks. There is a possibility that the turning points of the trend adjusted index could shift arbitrarily which is an undesired consequence. For example, this shift in turning points could happen because adding the mean growth rate from the CEI could turn some small negative monthly growth rates in the LEI into positive changes. This is an example for the case where the trend of the CEI is greater than the trend of the LEI. An example for the opposite case, which is relevant for composite indexes of some countries, is that adding the mean growth rate from the CEI could turn some small positive

---

[3]Before 1989, the reverse trend adjustment used an average of the four coincident indicators. Between 1989 and 1993 BEA equated the trend in the composite indexes to the trend in GNP.

[4]The inverse standard deviations are made to sum to 1 in keeping with index methodology since 1993.

monthly growth rates in the LEI into negative changes (e.g. when this adjustment was applied to the Australian LEI, it added several extra cycles).

Instead of removing the old trend and adding a new trend, an alternative way of making this reverse trend adjustment is by multiplying the monthly growth rates of the index by a factor that scales the trend of the index up or down to match the target trend. The multiplicative factor is the ratio of the long term average growth in the CEI to the long term average growth in the LEI. This alternative method ensures that the up and down movements in the final index are the same as that of the unadjusted index because the sign of monthly growth rates is not affected if this method is used. However, if the factor is greater than 1, the amplitude of the movements in the adjusted index will increase. On the other hand, if the factor is less than 1, the amplitude of the adjusted index will decrease. This could potentially pose a problem in the selection of turning points via Bry-Boschan algorithm especially if there are two sets of cyclical peaks and troughs close together in time and magnitude. Suppose in the unadjusted index the first peak is close in level but lower than the second peak. After the adjustment, if the growth rates before the first peak are stronger than those before the second peak, the adjusted index could show that the first peak was higher than the second. This would lead the first peak to be picked up by the Bry-Boschan algorithm. Therefore, the visual inspection of program determined turning points by an analyst rather than ready acceptance of the algorithm results becomes even more important.

It is worth noting that when the LEI is interpreted on an ongoing basis in real time, analysts want to know whether a peak has occurred with as little delay as possible. For this reason, the multiplicative method of trend adjustment is more accurate (assuming the original turning points are accurate). Whether that peak is part of a double peak and which one is higher seems to be less important in real time, even though it would play an important role in evaluating an existing index during a benchmark or during the development of a new index.

8.5 Benchmarking the Composite Indexes

The long history of the composite index is not revised during the monthly updates of the indexes. Once a year, during annual benchmark revisions in January, revisions that go back more than one year are incorporated into the history of the index. As part of this benchmark revision, the standardization factors described above are updated. Regular monthly updates to the leading, coincident, and lagging indexes, however, incorporate revisions to data over the past six months. This avoids potentially large changes in the indexes at the end of the year due to revisions to the component data which occur routinely throughout the year. The overall effects are minor because their composition – leading, coincident, or lagging indicators included in their respective indexes – are not changed. As long as the data revisions and updates are minor, benchmark revisions have a small impact on the indexes. The cyclical performance of all three indexes in the current cycle should not be visibly affected by benchmark revisions. At this time, the properties of each component and the composite indexes are reviewed.

It is rare that this review process uncovers a significant change in the behavior of one or more of the components. In that case, this issue is handled as part of a more comprehensive evaluation of the composite indexes discussed in the next section.

8.5.1 Reasons for Changes in the Composition of the Indexes

The long history of the composite indexes has been punctuated from time to time by changes in their composition as individual indicators were added, subtracted and their technical properties changed.[5]

---

[5] For more detailed discussion on changes in composition and their effect on the forecasting performance of the LEI see McGuckin and Ozyildirim (2004).

The first list of leading indicators dates back to Mitchell and Burns (1938). This original list of leading indicators later underwent major revisions six times; these revisions are documented in Moore (1950), Moore (1961), Moore and Shiskin (1967), Zarnowitz and Boschan (1975), Hertzberg and Beckman (1989), and The Conference Board (2001). The U.S. Commerce Department also made other changes on a smaller scale from time to time such as the removal of a component in 1987.

The annual benchmark revisions which update the history of the indexes to reflect revisions to the underlying data provide a good opportunity to review the performance of the indexes as well their components. The following sub-section lists basic principles or criteria for making changes in the index. Of course, each new component and the composite index itself are also subjected to the basic six criteria discussed above (i.e. Conformity, Consistent Timing, Economic Significance, Statistical Adequacy, Smoothness, Currency or Timeliness.)

8.5.2 Criteria based on quality of component indicators

The majority of these revisions in the components reflect improvements in the quality of statistics available at the time.[6] As new series of higher quality (better timeliness, more coverage, better statistical adequacy etc.) that measured the same concepts became available, they were substituted for the original series. On a few occasions, series have been discontinued by the source agency, and — unless good substitutes were available — they have been dropped from the list (e.g. index of net business formation discontinued in the 1970s).

There are also occasional efforts to improve the index by removing or replacing a component that became unduly volatile or otherwise deteriorated. A good example is the index of sensitive materials prices that was included in LEI prior to the 1996 revision by The Conference Board. This change in the composition of the LEI was a result of an evaluation of the index components by the criteria discussed in the next part. In general, such changes are infrequent as a large body of evidence over a long enough period of time is needed to make such a far-reaching change.

Producers of cyclical indicators must be sensitive to the effects of real changes in the economy's structure, institutions, and policies. Thus, new series have been added when accumulating evidence suggested strong cyclical performance ahead of recessions. A good example is the addition of real money supply to LEI in the 1973-75 review and benchmarking. Another is the addition of the yield spread by The Conference Board in 1996. This reflected research on the usefulness of this variable in forecasting and the recognition of the growing importance of Fed policy operations that worked through changes in short-term interest rates.

To summarize, the composition of an index can be revised because

- Existing indicators are discontinued or undergo methodological/definitional changes that make them less suitable

- New indicators become available

- Existing indicators deteriorate

- Structural change in the economy warrants new indicators

8.5.3 Criteria based on how the LEI is used or interpreted

According to those involved in the operations, the NBER and the U.S. Bureau of Economic Analysis (BEA) conducted no formal fitting exercises to improve forecasting with LEI. During the last review of the composite

---

[6]For comprehensive information on the structure and evolution of cyclical indicators and composite indexes, see Klein (1999a),Klein (1999b) and Zarnowitz (1992) chapters 10 and 11.

indexes in 1996, The Conference Board compared the new indexes constructed in 1996 with 10 components to those using the 11 indicators they inherited from the Bureau of Economic Analysis.[7] The Conference Board also examined the relationships of the leading index with the coincident index and RGDP on a historical basis. But the focus of these examinations is more on turning points and false signals — times when the LEI indicated a recession and only a slowdown materialized.[8]

While there is no evidence that fitting exercises were a key factor in the adoption of the new 1996 components, the new list of components did eliminate two false signals in the index. Surely, this would tend to improve the historical fit in the sense that movements in the LEI would more closely correlate with those in RGDP and the CCI, which are highly correlated themselves. Moreover, since new indicators are often closely related to the series they replace, the possibility of selection bias and overfitting issues cannot be dismissed out of hand (Sims (1994), where he makes this point in a review of Zarnowitz (1992)). This will be more important as the trend adjustment is added. Thus, a key issue remains: Does the LEI improve forecasts and how does it work in real-time out-of-sample tests?

If the following conditions are satisfied, a change in the composition of the index is warranted. All of these conditions point to an improvement of the index performance.

(a) The index has fewer false signals (extra cycles) historically when a new component is included.

(b) The inclusion of a component enables an index to track a cycle that is otherwise missed.

(c) A significant improvement in the cyclical timing pattern, in terms of its consistency and conformity, of the new index after the inclusion of the new component.

(d) The index is smoother with better articulated turning points.

(e) There is an improvement in the correlation of the cyclical behavior of the new index to the cyclical behavior of aggregate economic activity after the inclusion of the new component.

Conditions (a) through (d) are part of turning point analysis, and they reflect the historical performance of the index. An improvement with regard to these conditions is likely to improve the historical fit of a coincident index to reference cycle dates. Similarly, such improvements in a leading index make one more confident about its leading characteristics. However, they do not constitute direct evidence of the forecasting ability of the leading index. In order to show that a leading index is a better forecasting tool the following conditions need to be satisfied.

(f) Better real-time out-of-sample forecasting performance after the inclusion of a new component.

(g) Better real-time turning point forecasting using nonlinear models such as probit recession probability and/or other regime switching models.

These types of tests use the as-published versions of the index, helping to increase confidence in the real time forecasting ability. These conditions focus on the ex-ante performance of the index which does not include data revisions or composition changes in the index that would not have been available to a forecaster at the time of the forecast.

Note that better real-time, out-of-sample exercises in forecasting provide a necessary but not sufficient condition for reliable forecasts of future turning points. This is especially true if the economy goes through a structural break. Identification of structural breaks, which are rare, requires a considerable amount of data and much debate. Thus the leading index may be negatively affected by the failure to include or exclude an important component. However, it is worth noting that this negative effect is not likely to be large because the strength of the leading index lies in its ability to summarize many different aspects of the aggregate economy

---

[7]See the discussion in The Conference Board (2001) on pages 58-60.

[8]In fact, the new and old indexes did not show any material differences with respect to NBER recession dates. The new index has shorter leads at the 1969, 1981, 1990 business cycle peaks, but longer leads at the 1961, 1970, 1975, and 1980 business cycle troughs. For the six recessions since 1959, the average lead at peaks fell from 12 to 9 months and the average lead at troughs rose from 3 to 4 months. Differences in lead times of 1-2 months are generally considered insignificant if they appear to be random.

activity without reliance on any single indicator. Thus, if a cyclical movement is underway, it is likely to be signaled by a leading index with well selected components. Accounting for structural breaks can only improve the leading characteristics of the index further.

8.5.4 2012 Comprehensive Revisions to the LEI for the United States

Prior to the comprehensive revision of 2012 which introduced new components to the LEI, the last time the index had comprehensive revisions was in 1996, after The Conference Board received the responsibility for the LEI and the Business Cycle Indicators program from the Bureau of Economic Analysis at the U.S. Department of Commerce.

Extensive reevaluation of existing components of The Conference Board Leading Economic Index® for the United States based on the principles described earlier resulted in major changes in its composition in 2012. Following discussions with the Business Cycle Indicators Advisory Panel and other experts, The Conference Board decided to replace three of the ten components and make a minor adjustment to another component. The composition changes reflected in the new LEI address structural changes that have occurred in the U.S. economy in the last several decades. These changes to the index composition include:

1. Incorporating the new Leading Credit Index™(LCI) and omitting the real money supply (M2) component starting in 1990 (real M2 remains in the index before 1990);

2. Replacing the ISM Supplier Delivery Index with the ISM New Orders Index;

3. Replacing the Reuters/University of Michigan Consumer Expectations Index with an equally weighted average of consumer expectations of business and economic conditions using questions from Surveys of Consumers conducted by Reuters/University of Michigan and Consumer Confidence Survey by The Conference Board (after 1978, Reuters/University of Michigan Consumer Expectations Index remains in the index before 1978); and

4. Replacing "New Orders for (nondefense) Capital Goods" with "New Orders for (non-defense) Capital Goods excluding Aircraft".

Research by The Conference Board has shown that the new components perform better as leading indicators than the old components that were replaced, and that the new composition of the index should lead to its better performance, also in real time.[9] The new composition also conforms better to the structural changes the US economy has undergone in the last three decades. Based on its performance since 1990, and especially before and during the 2008-2009 recession, the new LEI should provide more accurate predictions of business cycle peaks and troughs.

The composite indexes and their components constitute an early warning system to analyze and predict real-time developments in national and regional economies as they move through business cycles. As these economies evolve through structural changes, the system of composite indicators also needs to change and adapt through comprehensive revisions such as the 2012 revisions to the LEI for the United States. The criteria used in the compilation and evaluation of the composite indexes also provide guidelines for their on-going maintenance and evolution as the overall economy transforms.

---

[9]See Levanon et al. (2015) and Levanon et al. (2012).

Annex

8.A Index Calculation Methodology

The outline of the basic procedure for constructing composite indexes is as follows (see The Conference Board (2001) and The Conference Board web site for further details, updates, and discussion):

1. Calculate month-to-month changes for each component, X_{it}, where $i = 1, \ldots, n$.

   The monthly change in each component of a composite index is calculated using a symmetric percent change formula. However, using percent changes or log differences makes little difference to the resulting index.

   If the component is not in percent change form, a symmetric alternative to the conventional percent change formula is used: $r_t = 200 * (X_t - X_{t-1})/(X_t + X_{t-1})$.

   If the component X is in percent change form or an interest rate, simple arithmetic differences are calculated: $r_t = X_t X_{t-1}$.

   If the component X is a diffusion index or an interest rate spread, the monthly level is used as the monthly change in the component, $r_t = X_t$. (Diffusion indexes are first normalized by subtracting their sample mean and dividing by their standard deviation).

   If a component is counter-cyclical, its month-to-month change is multiplied by (-1) to invert it before it contributes to the composite index.

2. Adjust the month-to-month changes by multiplying each series by the component's standardization factor, w_i. The standardization factors are calculated from the inverse standard deviations of the month-to-month changes calculated in step one. The inverse standard deviations for the components are normalized to sum to one. The results of this step are the monthly contributions of each component $c_{i,t} = w_i * r_{i,t}$.

3. Adjust the month-to-month changes by multiplying each series by the component's standardization factor, w. The standardization factors are calculated from the inverse standard deviations of the month-to-month changes calculated in step one. The inverse standard deviations for the components are normalized to sum to one. The results of this step are the monthly contributions of each component $c(t) = w * r(t)$.

4. Add the adjusted month-to-month changes across the components for each month to obtain the growth rate. This step results in the sum of the adjusted contributions, $s_t = \Sigma_{i=1}^n c_{i,t}$, or $s_t = sum(1 \text{ to } n)c(t)$ where n is the number of components.

   Coincident indexes are calculated by following the next step below. Leading and lagging indexes involve an addition step at this point to equalize their long term trends with that of the coincident index. For this trend adjustment, the sample mean of the sum of the adjusted contributions in the previous step is subtracted from the series, s_t, and the corresponding average growth rate from the coincident index is added. Thus, s_t now becomes the trend adjusted sum of contributions.

5. Compute preliminary levels of the index using the symmetric percent change formula listed in step (1). The index is calculated recursively starting from an initial value of 100 for the first month of the sample period (i.e. January 1987). The first month's value is $I(1) = 100$. The second month's value

$I_2 = I_1 * \frac{(200+S_2)}{(200-S_2)} = 100 * \frac{(200+S_2)}{(200-S_2)}$ or $I(2) = I(1) * (200 + s(2))/(200 - s(2))$ and this formula is used recursively to compute the index levels for each month that data are available.

6. The index is rebased to average 100 in the base year (e.g. 2010). The history of the index is multiplied by 100 and divided by the average for the twelve months of the base year (e.g. 2010).

8.B United States Composite Economic Indexes: Components

The Conference Board Leading Economic Index®

1. Average weekly hours (manufacturing)

2. Average weekly initial claims for unemployment insurance

3. Manufacturers' new orders, consumer goods and materials

4. ISM®new orders index (Institute for Supply Management)

5. Manufacturers' new orders, nondefense capital goods excl. aircraft

6. Building permits, new private housing units

7. Stock prices, 500 common stocks (S&P 500 Index)

8. Leading Credit Index™

9. Interest rate spread, 10-year Treasury bonds less federal funds

10. Avg. consumer expectations for business conditions (University of Michigan, The Conference Board)

The Conference Board Coincident Economic Index®

1. Employees on nonagricultural payrolls

2. Personal income less transfer payments

3. Industrial production

4. Manufacturing and trade sales

The Conference Board Lagging Economic Index®

1. Inventories to sales ratio, manufacturing and trade

2. Average duration of unemployment

3. Consumer installment credit outstanding to personal income ratio

4. Commercial and industrial loans

5. Average prime rate

6. Labor cost per unit of output, manufacturing

7. Consumer price index for services

Bibliography

Bry G. and Boschan C. (1971). Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, NBER Technical Paper N 20.

Burns A.F. and Mitchell W.C. (1946). Measuring Business Cycles, (National Bureau of Economic Research, New York).

Green G.R. and Beckman B.A. (1993). Business cycle indicators: Upcoming revisions of composite leading indicators, Survey of Current Business, October, Vol. 73, pp. 44-51.

Hertzberg M.P. and Beckman B.A. (1989). Business Cycle Indicators: Revised Composite Indexes, Survey of Current Business 72 (6), pp. 23-28.

Hodrick R.J. and Prescott E.C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation, Journal of Money Credit and Banking, 29(1), 1-16.

Klein P.A. (1999). The Leading Indicators in Historical Perspective, Business Cycle Indicators, 4(10), 3-4.

Klein P.A. (1999). The Leading Indicators in Historical Perspective, Business Cycle Indicators, 4(11), 3-4.

Klein P.A. and Moore G.H. (1985). Monitoring Growth Cycles in Market-Oriented Countries: Developing and Using International Economic Indicators, (Ballinger for NBER, Cambridge, Mass.).

Levanon G., Manini J.C., Ozyildirim A., Schaitkin B. and Tanchua J. (2015). Using Financial Indicators to Predict Turning Points in the Business Cycle: The Case of the Leading Economic Index for the United States, International Journal of Forecasting, 31 (2015), pp. 426-445.

Levanon G., Ozyildirim A., Schaitkin B. and Zabinska J. (2012). Comprehensive Benchmark Revisions of The Conference Board Leading Economic Index for the United States, in JSM Proceedings, Business and Economic Statistics Section. Alexandria, VA: American Statistical Association. 1127-1141.

McGuckin R.H. and Ozyildirim A. (2004). Real-Time Tests of the Leading Economic Index: Do Changes in the Index Composition Matter? Journal of Business Cycle Measurement and Analysis, OECD Publishing, CIRET, vol. 2004(2), pages 171-191.

Mintz I. (1969). Dating Postwar Business Cycles: Methods and Their Application to Western Germany, 1950-1967, Occasional Paper No. 107, (National Bureau of Economic Research, New York).

Mintz I. (1972). Dating American Growth Cycles, in Zarnowitz, V., ed., The Business Cycle Today, (National Bureau of Economic Research, New York).

Mitchell W.C. (1951). What Happens During Business Cycles: A progress report, (National Bureau of Economic Research, New York).

Mitchell W.C. and Burns A.F. (1938). Statistical Indicators of Cyclical Revivals, National Bureau of Economic Research, Bulletin 69.

Moore G.H. and Shiskin J. (1967). Indicators of business expansions and contractions, NBER Occasional Paper no. 103.

Moore G.H. (1950). Statistical Indicators of Cyclical Revivals and Recessions, New York: NBER.

Moore G.H. (1961). ed. Business Cycle Indicators, 2 vols., Princeton: NJ: Princeton University Press published for NBER.

Ravn M.O. and Uhlig H. (2001). On Adjusting the HP-Filter for the Frequency of Observations. CEPR Discussion Papers 2858, C.E.P.R. Discussion Papers.

Shiskin J. (1960). How Accurate?, The American Statistician, Vol. 14, No. 4, pp. 15-17

Sims C. (1994). Review of:Business Cycles: Theory, History, Indicators, and Forecasting by Victor Zarnowitz (1992) Journal of Economic Literature, 32(4), 1885-88.

The Conference Board (2001). Business Cycle Indicators Handbook, New York, NY.

The Conference Board (2013). Understanding Business Cycles: The Indicators Approach to Forecasting for Agility, ed. Ataman Ozyildirim, Research Report R-1519-13-RR

Zarnowitz V. (1972). The Business Cycle Today, National Bureau of Economic Research, New York.

Zarnowitz V. (1992). Business cycles: Theory, history, indicators and forecasting. The University of Chicago press. In Studies in business cycles, vol 27, ISBN 0-226-97890-7.

Zarnowitz V. and Boschan C. (1975). Cyclical Indicators: An Evaluation and New Leading Indexes, Business Conditions Digest, May 1975.

9

Alternative Detrending Methods

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

James Mitchell
Warwick Business School, University of Warwick
james.mitchell@wbs.ac.uk

Table of Contents

9.1 Introduction: The importance of detrending time-series

Time series are often decomposed into trend and cyclical components by statisticians, and other applied users particularly economists, in order to facilitate inference. This is because many time-series increase - trend upwards - over time; they are "nonstationary" (defined heuristically), in the sense that they lack a constant mean to which they revert. Macroeconomists, in particular, therefore often detrend data in order to understand whether changes to the economy are temporary or permanent, that is whether they represent cyclical or trend movements. Policymakers assess the degree to which the economy is operating above or below its trend level. But the importance of the "cycle" in popular debate can tend to run ahead of the problems in measuring it; trends and cycles are not directly observable. Their measurement is contingent on specification of some model, whether formalised or not, which seeks to identify and estimate trend and cycle.

Moreover, since trends and cycles can interact and influence each other there is considerable diversity and uncertainty about the model used to undertake the decomposition. The nature of the trend component is disputed; different trend models can be distinguished according to whether the trend is constant and predictable over time ("deterministic") or changing and unpredictable ("stochastic"). In turn, one can model both deterministic and stochastic trend components in different ways. There is no unique way of modelling these trend components. In turn, the debate over whether the trend is stochastic or deterministic has important economic implications. If the trend is deterministic, so all fluctuations in economic activity are associated with the cycle, then there is scope for macroeconomic stabilisation policy in a way. There is not if the trend is itself changing and responsible for fluctuations in economic activity.

In this Chapter we provide a review of the large detrending literature that has therefore arisen. We seek to classify different modelling approaches and discuss their relative advantages. The choice of what measure, or estimator, of the cycle to use is more than a dry academic issue. As the analysis of Canova (1998), for example, showed inference can be sensitive to measurement.

As a specific illustrative example, we focus on detrending GDP, where the cyclical estimate has its own name - the "output gap". But this does not affect the generality of our discussion, beyond the fact that many of the multivariate and structural decomposition methods which we consider (and define below) are tailored to detrend GDP to the extent that they look to economic theory for guidance on what form the model should take.

Indeed, one of the distinguishing features of alternative detrending methods is the degree to which they rely on the data (statistical evidence) versus economic theory to define the model. We will expand on this below and explain that there can be a tension between statistical evidence (what the data "say") and economic theory. Resolving this, and defining the relative importance of economic theory and statistical evidence, varies across detrending models and remains the subject of lively debate. This is so particularly in the face of an evolving macro-economy, with time-series subject to possible structural breaks, due to events like the global financial crisis and/or changes in productivity growth, and/or changes in not just their conditional means but higher moments too.

The plan of the remainder of this chapter is as follows: Section 9.2 distinguishes growth from classical cycles. Section 9.3 classifies alternative detrending methods; with a view to drawing out the pros and cons of different methods. But throughout we will stress that the relative attractiveness of different detrending methods is largely subjective. Section 9.4 then takes this classification up in more detail and distinguishes parametric from nonparametric detrending methods, considering a range of alternative methods from each camp. Section 9.4.3 then explains how the Beveridge-Nelson decomposition can be viewed as a general, limiting detrending method. Section 9.5.1 continues the classification, distinguishing univariate (or non-structural) from multivariate (or more structural) methods; to illustrate the wide range of available multivariate detrending methods seven types of multivariate models are examined. Section 9.6 delineates linear and nonlinear models. Given the wide variety of alternative detrending methods, Section 9.7 reviews methods to accommodate uncertainty

over the preferred method. Section 9.8 considers how real-time or end-of-sample cyclical estimates are often uncertain. Section 9.9 examines how competing cyclical estimates might be evaluated. Finally, Section 9.10 provides an illustrative application detrending Euro-area GDP in real-time. In so doing it illustrates uncertainties associated both with real-time cyclical estimates and across alternative detrending methods; and emphasises that, in general, there is no single detrending method that will consistently dominate competitors across alternative evaluation criteria.

Detrending techniques presented in this chapter are supposed to play a crucial role when constructing a cyclical indicator having as a reference cycle the growth cycle (i.e. output gap). The composit leading indicators produced by the OECD (see chapter 10) is the most well known indicator of such type.

9.2 Distinguishing the "growth" and "classical" business cycles

Detrending methods seek, in the following stylised way, to decompose a time series y_t into a trend component, μ_t, a cyclical component, c_t, and an irregular shock, ε_t (which may also capture seasonal and calendar components when y_t is assumed to be seasonally unadjusted rather than adjusted data):

$$y_t = \mu_t + c_t + \varepsilon_t \dots \tag{9.1}$$

where t $(t = 1, \dots, T)$ denotes time. As we review below, both statisticians and economists have employed various algorithms to identify and estimate the cyclical component, c_t. Economists, in particular, when detrending macroeconomic variables, like GDP, call c_t the "growth cycle" or the "deviation cycle".

National statisticians call $\mu_t + c_t$ the "trend-cycle". In practice, the trend-cycle is calculated by removing the seasonal, calendar and irregular components (i.e. ε_t) from the original non-adjusted data. Data in this form are published by some national statistical offices and bodies like the OECD.

The trend-cycle represents the underlying general direction of an unadjusted data series. It combines long-term (trend) and medium-to-long-term (cycle) movements. For the purposes of seasonal adjustment, both the long-term trend and medium-term cycles are treated as the trend component of a time series.

Detrending methods also differ according to whether they are best applied to seasonally adjusted or unadjusted data. As we review below, some detrending methods explicitly seek to eliminate any seasonal component to the time series, ε_t, and can therefore be applied to seasonally unadjusted (raw) data; while others would effectively place ε_t in the cycle. In these cases, to avoid "excessive" noise in the composite seasonal component, $c_t + \varepsilon_t$, it can be prudent to apply the detrending method to seasonally adjusted data, where ε_t has been removed at an initial step. National Statistical Offices, of course, often publish seasonally adjusted data themselves.

As discussed in detail by Zarnowitz and Ozyildirim (2006) it is important to distinguish this "growth" measure of the cycle from the "classical" business cycle. The growth cycle is dependent on the application of some detrending method, whereas the classical cycle is not. The classical cycle is a sequence of expansions and contractions in the (perhaps logarithmic) levels of a large array of series, representing the levels of total output, employment, and many other components and related processes. These expansions and contractions in the level of these series can be analysed without trend adjustments estimated following the form in (9.1). A dating rule is used to classify periods of expansion from periods of contraction; e.g., see Harding and Pagan (2002) and Bry and Boschan (1971). This approach to defining the "classical cycle" is associated with the National Bureau of Economic Research (NBER) and their dating of business cycle turning points, an approach also followed by The Conference Board; e.g. see Stock and Watson (2010).

9.3 Classifying detrending methods

9.3.1 Trend-cycle decompositions methods

Trend-cycle decompositions methods, that give rise to the so-called growth cycle, can be usefully grouped into the following (overlapping) categories:

1. parametric and non-parametric detrending algorithms;

2. univariate (and often non-structural) and multivariate (often more structural) approaches;

3. linear and nonlinear detrending models.

The second group of trend-cycle decompositions commonly delineates detrending methods that are purely empirical and statistical from methods where economic theory, to varying degrees, is used to identify and estimate the trend and cyclical components. At the extreme, economic theory gets most of the weight. But, as we indicate below, the distinction between these alternative approaches is often blurred.

Table 9.1 provides an initial attempt to classify further the alternative detrending methods we consider below. Our classification choices are explained in the text below, when we review the different detrending methods in more detail.

The classification criteria in Table 9.1 are whether the method is (principally) parametric or not, multivariate, seeks to maximise statistical "fit" as measured via one metric or other and whether the detrending method has a high economic content, which facilitates economic interpretation. Finally, we indicate whether the method is easy to compute (using freely available software).

There can but need not be trade-offs between these criteria; e.g. between a method that seeks to maximise statistical fit and economic content. As it is not always clear cut how to classify a given detrending method, a "?" in Table 9.1 indicates when the classification is either subject to particular uncertainty or when classification could be either way.

Table 9.1 emphasises from the outset that the different detrending methods have different properties. These can be seen as the basis for normative statements about the relative pros and cons of the different methods. But, as indicated, clearly it is in the eye of the beholder whether, for example, more economic content is a pro or a con. On this basis, our view is that there is effectively "model uncertainty" about the preferred way to detrend a time-series - a topic we return to in section 9.7. It is difficult, if not impossible, to agree on a single preferred detrending method. It therefore seems prudent to consider various methods; and assess the robustness of inference to the chosen detrending method.

For completeness we should also mention there are "direct" measures of the cycle, in particular of the output gap. The "direct" approach typically relies on a qualitative, rather than quantitative, survey to measure the degree of capacity utilisation in the economy. Our focus in this chapter is on detrending using quantitative data.

A related classification of alternative detrending methods is offered by Chagny and Döpke (2001) who differentiate four main groups:

1. *non-structural approaches*;

2. *direct measures of the cycle from survey data*;

3. *structural approaches* and

4. *multivariate approaches*.

Our classification above does not distinguish the multivariate approach from the structural approach. This is as, although it is certainly true that some multivariate detrending methods rely more on economic theory than

Table 9.1: Classifying alternative detrending methods

	Parametric	Multivariate	Statistical fit	Economic content	Ease of computation and interpretation
Linear	✓	×	?	×	✓
Beveridge-Nelson	✓	?	✓	?	?
Unobserved components	✓	?	✓	?	?
Moving average	×	?	?	×	✓
Hodrick-Prescott	×	?	×	×	✓
Band pass	×	?	?	✓	✓
Phase Average Trend	×	×	×	?	?
Structural VAR	✓	✓	?	✓	?
Production function	✓	✓	×	✓	×
DSGE	✓	✓	×	✓	×
Regime switching	✓	?	?	✓	×

Notes to Table: a "?" indicates when the classification is either subject to particular uncertainty or when classification could be either way, perhaps depending on how the method is implemented.

others, all are reliant to some extent as they all exploit dependencies between data when seeking to extract trend components.

In the following sections we consider each of these three approaches and each of the detrending methods considered in Table 9.1 in turn. Before doing so, we distinguish deterministic from stochastic trends.

9.3.2 Deterministic versus stochastic trends

At the most basic level different detrending methods can be distinguished according to whether they assume y_t has a stochastic or a deterministic trend. When y_t has a deterministic trend any shocks to the series are temporary — and ultimately the series reverts to trend. This means that one can forecast distant values of y_t with accuracy since in the long-run y_t will have reverted to its predictable trend. In contrast, when y_t has a stochastic trend shocks persist. This means that it becomes increasingly hard to forecast future values of y_t. Forecast uncertainty increases with the forecast horizon. Time-series with a stochastic trend can be rendered stationary by differencing (e.g. $y_t - y_{t-1}$), i.e. by performing the modelling and analysis on period to period changes.

Nelson and Plosser (1982) considered the empirical evidence as to whether macroeconomic time-series have deterministic or stochastic trends. However, it is probably fair to say that to-date the evidence remains mixed. This is particularly so since it is hard to distinguish series with stochastic trends from series with deterministic trends, but subject to occasional breaks in the nature of this trend; see Perron (1989). Nelson and Kang (1981) and Cogley (2001) considered the econometric consequences of misspecification of the nature of the trend component.

Below when we review the different detrending methods we note the different assumptions they make about deterministic versus stochastic trends; and/or how flexible they are to this assumption. Given that statistical tests to distinguish deterministic from stochastic trends are known to be subject to low power, it can be an

advantage to use a detrending method which does not take a strong stance on the nature of the trend — and can accommodate either extreme. This way one is less likely to apply the "wrong" detrending method to the series of interest.

9.4 Parametric versus nonparametric trend-cycle de-compositions

Before considering some popular parametric and nonparametric methods, we should emphasise at the outset that the apparent distinction between parametric and nonparametric detrending methods, that reflects a long standing methodological debate, is less acute than often appreciated.

As we shall indicate below, many nonparametric methods can be rationalised as parametric ones. In particular, the Hodrick-Prescott and ideal band pass filters (such as the Baxter-King filter) can be motivated within an unobserved components framework; see Harvey and Jaeger (1993), Harvey and Trimbur (2003) and for the multivariate case Valle e Azevedo (2006) and Creal et al. (2010). Furthermore, the parametric methods, like the nonparametric ones, are "simply" taking weighted averages of the data in the sense that they involve applying a filter, given by the generic lag polynomial, $a(L)$, to y_t, the series, so that the cyclical component is derived as $c_t = a(L)y_t$. Harvey and Koopman (2000) provide methods to derive these weights, given by $a(L)$, for specific parametric detrending methods.

Despite this relationship, the nonparametric methods are often seen as less reliant on a statistical model. As a result they are viewed as numerical, more mechanical procedures motivated by convenience and transparency. The nonparametric filters employed to extract c_t tend to be examined within the frequency domain. Their detractors argue that since these nonparametric filters can be seen through the lens of a statistical model, one should use the data to tune the filter and thereby establish the statistical properties of the filter. It is preferable in their view not to impose restrictions on the nature of the trend component *ex ante*, as is common with nonparametric measures.

9.4.1 Parametric measures

Linear detrending

The simplest hypothesis which can be formulated for the representation of the trend component, μ_t, is that it follows a linear deterministic trend. But the practical relevance of such methods for the estimation of μ_t is perhaps quite low, since it does seem to fair to say that there is a consensus that most economic time-series are not perfectly explained simply by a deterministic trend. However, as mentioned above, it remains an open question whether many time-series are well represented by a sequence of deterministic trends subject to breaks — "segmented" trends — which we consider below.

The trend component, μ_t, is assumed to be a linear function of the time so that $\mu_t = \beta_0 + \beta_1 t$. Given equation (9.1) we then see that the cyclical component is the residual from a regression of y_t on a constant, β_0, and time, t:

$$y_t = \beta_0 + \beta_1 t + c_t.$$

It is therefore apparent that c_t will include irregular shocks, ε_t. There is, therefore, a case for detrending using seasonally adjusted data on y_t, so that ε_t does not also include any seasonal component to y_t.

When y_t is in fact a logarithmic transformation of the original time-series series the estimated coefficient, $\hat{\beta}_1$ is the average growth of y_t over the T periods.

Linear detrending with breaks

The assumption that there is a single constant trend to which the time-series of interest reverts has been questioned in many economic applications. Perron and Wada (2009), to focus on a relatively recent example, argue that the trend component of US real GDP has been subject to breaks. In particular, they argue that there was a change in the slope of the trend function around 1973, the time of the global oil price shocks.

As well as segmented deterministic trends of this type, which accommodate infrequent shifts to the trend, continuous but importantly perfectly predictable change in the trend can be modelled by a quadratic trend or other such polynomials.

But a simple deterministic time trend can only imperfectly approximate stochastic nonstationarity even if it is allowed to be subject to structural shifts. A stochastic trend, to which we now turn, can be seen as the limiting case of allowing permanent shocks each period, not at discrete (finite, selected) points in time.

Beveridge-Nelson (BN) decomposition

The BN decomposition, as suggested by Beveridge and Nelson (1981), is one of the most prominent procedures for finding the cyclical component of a nonstationary (due to a stochastic trend) time series. Beveridge and Nelson originally phrased their procedure in terms of forecasts of an ARIMA model. But a state-space interpretation can also be provided (see Morley (2002)). Following this link and drawing on Watson (1986), Morley et al. (2003) have recently established that, in fact, the trend coming out of the BN decomposition and that resulting from Unobserved Components models (considered below) are identical, under certain conditions, as we explain below.

Beveridge and Nelson (1981) defined the trend component as the limiting forecast of y_t adjusted for its mean growth rate. Specifically, to put this mathematically, the Beveridge-Nelson trend is defined as:

$$\mu_t = \lim_{h \to \infty} [E(y_{t+h}|\Omega_t) - h\tau]$$

where h is the forecast horizon, Ω_t is the information available when the forecast for y_t h periods ahead is made. $\tau = E(\Delta y_t)$ is the mean of the first difference of y_t, the average growth rate when y_t is subject to the logarithmic transformation. The resulting cyclical component is $c_t = \mu_t - y_t$.

To compute μ_t and c_t the traditional approach first estimates an ARMA model in first differences, Δy_t. This is then used to produce the forecasts

$$c_t = \lim_{h \to \infty} \sum_{j=0}^{h-1} E(\Delta y_{t+h-j}|\Omega) - h\tau$$

on the basis of which the cyclical component is then computed.

In practice, of course, the forecast horizon h must be fixed. A series of papers have suggested computationally efficient and exact procedures to compute these forecast. An elegant solution is to cast the ARMA in state-space form, as seen in Morley (2002).

To add more detail, to motivate the BN decomposition, consider an $ARIMA(p, 1, q)$ process:

$$a(L)\Delta y_t = f + b(L)e_t$$

where f is a constant and $a(L) = \sum_{j=0}^{p} a_j L^j$ and $b(L) = \Sigma_{j=0}^{q} b_j L^j$ are two lag polynomials of order p and q, respectively. Inverting $a(L)$ the model can be rewritten as an infinite MA process:

$$\Delta y_t = h + c(L)e_t \tag{9.1}$$

where $h = f/\Sigma_{j=0}^{p} a_j$ and $c(L) = b(L)/a(L)$. The right term in equation (9.1) can be decomposed as:

$$\Delta y_t = h + c(1)e_t + [c(L) - c(1)]e_t$$

which can be written as:

$$\Delta y_t = h + c(1)e_t + (1 - L)c^*(L)e_t$$

where $c^*(L)$ is the lag polynomial such that $(1 - L)c^*(L) = c(L) - c(1)$. Pre-multiplying both sides of the equation above by $(1 - L)^{-1}$ and assuming $y_0 = e_0 = 0$ we see that:

$$y_t = ht + c(l) \sum_{j=1}^{t} e_j + c^*(L)e_t \tag{9.2}$$

Equation (9.2) gives the Beveridge and Nelson decomposition for the series, y_t. The permanent component is given by:

$$\mu_t = ht + c(1) \sum_{j=1}^{t} e_j$$

where ht is the "deterministic trend" an $c(1) \sum_{j=1}^{t} e_j$ the "stochastic trend". The trend component in the BN decomposition is thus seen to follow a random walk, since it can be rewritten as

$$\mu_t = \mu t - 1 + h + c(1)e_t$$

This random walk nature of the trend is often seen as a weakness but, as we explain in Section 9.4.3, in fact it follows from its forward-looking nature; indeed any trend component converges in expectation on the BN trend as the horizon increases (see Garratt et al. (2006)).

The transitory component is given by:

$$c_t = c^*(L)e_t$$

The innovations in the trend and in the cyclical components are both proportional to e_t and thus are perfectly (in fact negatively) correlated. In an ARMA model there is just one type of shock which has both permanent and transitory effects. The permanent effect is the change in the trend.

Empirically it is often found that the estimated trend component, μ_t, is highly irregular. We can appreciate this since μ_t involves cumulating shocks, e. Indeed the estimated trend component can be more volatile than the series, y_t, itself.

Unobserved components models

A popular approach to econometric modelling is the so called "structural time series" approach that phrases a model in terms of unobserved components (UC) that have a direct economic interpretation. In the univariate case, they usually take the general form of equation (9.1) with other components sometimes added in, often to accommodate seasonal or irregular movements.

UC models can be seen as a flexible data-driven means of detrending. They are flexible in the sense that they can accommodate both deterministic and/or stochastic non-stationarity.

A generic UC representation (see Harvey (1985), Harvey (1989)) takes the form:

$$
\begin{aligned}
y_t &= \mu_t + c_t + \varepsilon_t \\
\mu_t &= \mu_{t-1} + \tau + \eta_t^\mu, \eta_t^\mu \sim IID(0, \sigma_\mu^2) \\
&\quad c_t \text{ is stationary}, \eta_t^c \sim IID(0, \sigma_c^2)
\end{aligned}
$$

where ε_t, the irregular component, may be decomposed into seasonal components and calendar effects when modelling seasonally unadjusted data for y_t. When modelling seasonally adjusted data ε_t would typically be assumed to be a mean zero IID process.

The stochastic trend component in this stylised representation above is assumed to follow a random walk possibly with drift, τ. In many representations this drift term is itself time-varying. The process for the cyclical component is typically taken to be some ARMA or trigonometric process, with an innovation η_t^c. The flexibility of the UC approach can be easily appreciated by considering that when $\sigma_\mu^2 = 0$ the trend component reduces to a deterministic trend. This flexibility is a perceived advantage of UC methods; they can adapt to the time-series properties of the data of interest – and can accommodate both deterministic and stochastic trends. This means as long as the UC model applied is sufficiently general, one does not need to take an *a priori* stance on the time-series properties of the data and one can apply the model to any series.

In traditional implementations of the UC approach (e.g. Harvey (1985)) the innovations to the trend and cyclical components are assumed orthogonal, so that $E(\eta_t^\mu \eta_t^c) = 0$. Restrictions are also often imposed to ensure "smoothness" of the trend component by imposing $\sigma_\mu^2 = 0$ but letting the drift term evolve as $\tau_t = \tau_{t-1} + \eta_\tau, \eta_\tau \sim IID(0, \eta_t^2)$.

The parameters of the UC model are estimated either by maximum likelihood, exploiting the prediction error decomposition (see Harvey (1989)), or by Bayesian methods. Given these parameters, the Kalman filter and Kalman smoother provide estimates for the cyclical component c_t and the trend μ_t. The Kalman filter conditions on data known up to time t, while the smoother uses full-sample information by estimating the trend and cyclical components using data known up to time T. This means it provides more efficient historical estimates of the trend and cycle than the Kalman filter; however, as we consider below, in real-time (i.e. at the end of the sample, when $t = T$) the filtered and smoothed estimates are identical. See Harvey (1989) for additional details. Computation of UC cycles and trends does rely on maximisation algorithms, and is therefore not simply mechanical, but is facilitated by software like STAMP.

Since the UC model implies a restricted ARMA one can relate the UC trend to the aforementioned BN trend; see Morley et al. (2003) and Watson (1986). The UC, or Kalman filter, trend conditional on data through time t, $E(\mu_t | \Omega_t)$, will equal the BN trend when the UC model implies the same (reduced-form) ARMA model that the BN trend is based upon. Differences between the UC and BN trend arise because of the assumed restriction in the UC model that $E(\eta_t^\mu \eta_t^c) = 0$. When this zero correlation restriction is relaxed the UC and BN decompositions are identical. See also Proietti (2006) on univariate trend-cycle decompositions with correlated disturbances.

More complex detrending methods can also be considered within the UC framework, given its flexibility. In-

deed, for example, Perron and Wada (2009) show how their deterministic trend with breaks model can be proxied by an UC model where the errors are not assumed normal but allowed to be more flexible, specifically when they follow a mixture distribution. Trends and cyclical components in the UC model can also be modelled within a Bayesian framework; e.g. see Harvey et al. (2007). This enables prior notions about the duration of cycles to be imposed when detrending. Parameters in the UC model can also be allowed to vary over time to accommodate structural instabilities; this can include letting the conditional variance vary over time; e.g. see Stock and Watson (2007).

9.4.2 Non-parametric measures

Theoretical foundations

Non-parametric methods offer an alternative means of extracting trend and cyclical components from a time series. They do not explicitly rely on a parametric, or structural, model. Nevertheless, like the parametric methods, non-parametric methods extract a signal from a time series by taking a weighted average of the time series by means of an *ad hoc* filter. This weighted average can be either linear or nonlinear. Leaving an example of the non-linear case to Section 9.6, consider the linear case first. Then, as considered above, any nonparametric method of extracting a cyclical component, c_t, from a time series y_t takes the form:

$$c_t = a(L)y_t$$

where the filter $a(L)$ is expressed as the following polynomial

$$a(L) = a_{-r}L^{-r} + \ldots + a_{-1}L^{-1} + a_0 + a_1 L^1 \ldots + a_s L^s$$

The trend component of , y_t, μ_t is then given as

$$\mu_t = [1 - a(L)]y_t$$

Alternative nonparametric filters can be distinguished according to the properties of the lag polynomial, $a(L)$. The next five subsections discuss different linear filters.

Simple moving average filters

Moving average detrending has been widely used by both academics and policymakers for many years. A particular moving average filter that has been widely used is the centred moving average. Computed over an odd number of periods, such as $2m + 1$, the centred $MA(2m + 1)$ filter is defined such that

$$a_j = \frac{1}{2m + 1}, j = 0, \pm 1, \ldots, \pm m \tag{9.3}$$

For an even number of periods the centred $MA(2m)$ filter is defined by

$$a_j = \frac{1}{2m}, j = 0, \pm 1, \ldots, \pm(m - 1)$$
$$a_j = \frac{1}{4m}, j = \pm m \tag{9.4}$$

Following the work of Slutsky (1937), Osborn (1995) analysed the properties of the filter. In its simplest form the filter $a(L)$ is assumed: (i) symmetric, i.e. $a_k = a_{-k}$, $\forall k$ (ii) $r = s$, and (iii) the weights sum to zero $a(1) = 0$. Under these assumptions $a(L)$ can be factored such that

$$a(L) = (1 - L)(1 - L^{-1})d(L).$$

Moving average filters therefore contain two differencing operators and can render second order (quadratic) deterministic processes stationary and eliminate stochastic non-stationarity of two unit roots.

To analyse further the properties of these moving average filters it is useful to consider the filters in the frequency domain using the Fourier transformation. See Massmann et al. (2003) for a survey and more technical details. These methods let one characterise the effect of the filter $a(L)$ on y_t by examining the gain and the phase. The gain yields a measure of the increase (or the decrease) in the amplitude of the filtered series over the original series; and the phase yields a measure of the time displacement attributable to the linear filter.

Symmetric filters, such as (9.3) or (9.4), involve no phase shift, so the effect of a symmetric filter on y_t is solely captured by the gain. Importantly, the gain is dependent on the time-series properties of y_t, namely the number of unit roots in y_t; see Osborn (1995). Yule (1921) and Slutsky (1937) showed that moving average filters can generate spurious cycles, i.e. the gain can exhibit a cyclical pattern, particularly when y_t is a single unit root process. This potential weakness of moving averages explains why in Table 9.1 above, despite their statistical simplicity which is an attraction, we classify this detrending method as "?". The perceived attractiveness of (symmetric) moving average detrending methods is contingent on the underlying time-series properties of the data being detrended. Only when this is understood can one be confident in the performance of moving average detrending methods.Other types of moving average filter are also used widely. For example, the Henderson moving average is used by the U.S. Bureau of Census' X-11 and X-12 procedures to extract the cycle-trend component of a time series. It must be remarked that the longer is the filter the smoother is the resulting trend; so the choice of the filter length must depend on the purpose of the analysis

Differencing filters

First differencing is widely used to filter time-series. Predicated on the assumption that the trend component μ_t is a random walk with no drift, c_t is stationary and $E(\mu_t c_t) = 0$ an estimate of c_t is obtained as $(1 - L)y_t$. The properties of the first difference operator can again be analysed by looking at the gain and phase; the first difference operator attenuates the low frequencies and removes up to one unit root in y_t. Clearly if there is more than one unit root in the data the first differencing filter is not sufficient to render the series stationary; and if the underlying data were already stationary the filter will over-difference and thereby induce a moving average structure to the data, where there is a unit root in the moving average term. The first differencing filter is therefore very specific, in what it assumes about the time-series properties of y_t, and clearly not appropriate in general situations.

Hodrick Prescott (HP) filter

The Hodrick and Prescott (1980) filter is widely used in macroeconomics to fit a smooth curve through a set of points. HP minimise the variance of y_t subject to a penalty for variation in the second difference of μ_t, i.e.

$$\min_{\{\mu_t\}_{t=1}^T} \sum_{t=1}^{T} \left[(y_t - \mu_t)^2 + \lambda \left[(\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}) \right]^2 \right] \tag{9.5}$$

where λ controls the smoothness of the trend.

Therefore, in practice, users of the HP filter are required to choose one parameter, λ. For quarterly macroeconomic data λ is most commonly set equal to 1600. For monthly data Ravn and Uhlig (2002) suggest that λ should be set at 129600. But essentially these choices are *ad hoc*. By contrast, Pedersen (2001) derives an optimal estimator applicable when the modelled process is stationary. Macroeconomic data, however, are usually believed to contain a unit root so that the Pedersen estimator is not applicable. We can examine briefly the properties of the HP filter in more technical detail by seeing that the solution to (9.5) yields

$$a(L) = \frac{\lambda L^{-2}(1 - L)^4}{\lambda L^{-2}(1 - L)^4 + 1} \tag{9.6}$$

The HP filter is seen from (9.6) to contain four differencing operators. It therefore renders stationary any integrated process up to fourth order; see King and Rebelo (1993). When y_t is nonstationary one can analyse the properties of the HP filter by decomposing the filter into two operations: the first renders y_t stationary by an appropriate transform and the second operates on the resulting stationary series. When y_t is trend stationary applying the HP filter is equivalent to detrending y_t to make it stationary and then applying the HP filter to deviations from trend; see Cogley and Nason (1995). The effect of the filter again depends on the properties of y_t; see Harvey and Jaeger (1993), Cogley and Nason (1995) and Pedersen (2001). One may then analyse the properties of the filter in the frequency domain by examining the second operation; see Cogley and Nason (1995). This second operation is not a high pass filter; the gain has a peak at business cycle frequencies which is not present in y_t. HP filtering of nonstationary series can therefore generate spurious results. Pedersen (2001) argues that spurious *cycles* are not generated, in the sense that there is no cycle in the gain, just a peak.

Since the HP filter removes low frequency but not high frequency components, there is an argument for its use on seasonally adjusted data. Otherwise the estimated cyclical component may contain seasonal components too. Computation of the HP filter is straightforward as it is implemented in many software packages, such as EViews and STAMP.

Importantly, a parametric interpretation can be given to the HP filter. Harvey and Jaeger (1993) show how the HP filter may also be rationalised as the optimal estimator in a restricted unobserved components model.

Ideal Band Pass Filters and the Baxter-King and Christiano-Fitzgerald approximations

An ideal band pass filter is used to isolate the components of a time series that lie within a given range of frequencies, say ω_1 and ω_2. The period of the cycle (in quarters) is given as $p = 2\pi/\omega$.

Economic theory can play a role in defining these frequencies. In particular, when interest lies in extracting the periodic components of an economic time series that can be associated with the business cycle, the bands can be chosen consistent with priors about the duration of the business cycle. For example, it is widely believed that a business cycle lasts between 1.5 and 8 years; the lower band can then be set at 18 months and the upper band at 96 months. This removes low frequency trend variation and smooths high frequency irregular variation, while retaining the major features of business cycles. Band pass filters can therefore be applied to seasonally adjusted or unadjusted data, given that high frequency irregular variations are removed.

The removal of higher frequency components from the time series, such as seasonal components contained in ε_t in (9.1), means band pass filters can be applied to both seasonally adjusted and unadjusted data. Since the ideal band pass filter requires a moving average of infinite order, in practice an approximation is required. Two popular approximations are the Baxter and King (1999) (BK) filter, and secondly the Christiano and Fitzgerald (2003) (CF) filter. Both are readily computed using publicly available software (as a quick Google search reveals) or built into commercial software (such as EViews).

The BK filter is the difference between two ideal low-pass filters with cut-off frequencies ω_1 and ω_2. The ideal low-pass symmetric filter is of the form $b(L) = \Sigma_{h=-\infty}^{\infty} b_h L^h$ and Baxter and King (1999) show that the

weights in these ideal low-pass filters, with cut-off ω_j ($j = 1$ and 2), are $b_0 = \omega_j/\pi$, $bh = \sin(h\omega_j)/h\pi$.

The approximation proposed by Baxter and King (1999) has been shown to fail to filter out the desired components when y_t is nonstationary; see Murray (2003). In addition, end of sample estimates are unavailable since the Baxter-King approximation requires a number of lead and lag observations; i.e. $b(L)$ cannot in practice run from $h = -\infty$ to $h = -\infty$ and a moving average of length K loses $2K$ observations (K at the beginning of the sample and K at the end). Baxter and King note that there is no "best" value for the truncation, K, and K depends on the length of the sample under consideration. The larger is K the better is the approximation to the ideal filter, but the more data are lost at the beginning and end of the sample. In macroeconomic samples K is often set to 12; but this is not a rule. This means 3 years' worth of end-of-sample estimates are unavailable.

To avoid this end-point problem one might use an approximation to the band pass filter based on the difference between two Hodrick-Prescott filters with smoothness parameters:

$$\lambda_j = \frac{1}{4(1 - \cos(\omega_j)^2)} \; for \; (j = 1, 2). \tag{9.7}$$

So when interest lies with cycles between 5 and 32 quarters $\omega_1 = 0.196 \Leftrightarrow \lambda_1 = 677$ and $\omega_2 = 1.256 \Leftrightarrow \lambda_2 = 0.52$. Such a cyclical estimator is noisier than Baxter-King, but does provide end-of-sample estimates. The OECD, for example, refer to such filters - in which a large and then small value of λ is used (to detrend and then smooth, respectively) - as "double" Hodrick-Prescott filters; see Nilsson and Gyomai (2008).

Christiano and Fitzgerald (2003) explain that the optimal approximation to the ideal band pass filter requires knowledge of the true process for y_t. They derive the optimal approximation when y_t is integrated of order one (i.e. there is a single stochastic trend in the series of interest), a common benchmark in macroeconomics. The cycle then takes the form $c_t = b_0 y_t + b_1 y_{t+1} + \ldots + b_{T-1-t} y_{T-1} + \tilde{b}_{T-t} y_T + b_1 y_{t-1} + \ldots + b_{t-2} y_2 + \tilde{b}_{t-1} y_1$ where:

$$
\begin{aligned}
b_j &= \frac{\sin(jb) - \sin(ja)}{\pi j}; j \geq 1 \\
b_0 &= \frac{b - a}{\pi}, a = \frac{2\pi}{\lambda_1}, b = \frac{2\pi}{\lambda_2} \\
\tilde{b}_k &= -0.5 b_0 - \sum_{j=1}^{k-1} b_j
\end{aligned}
$$

where λ_1 and λ_2 are determined via equation 9.7.

Christiano and Fitzgerald (2003) find this particular approximation works well for standard macroeconomic time series. In contrast to the Baxter-King filter, by using an asymmetric rather than symmetric filter, Christiano and Fitzgerald (2003) use the whole time-series and can obtain end-of-sample estimates. This comes at the expense of phase shifts in the estimated cyclical components. In the long run their filter converges to the optimal filter. Again a parametric, model-based interpretation can be given to this filter; see Harvey and Jaeger (1993).

The Phase Average Trend (PAT)

The PAT, used widely including by the OECD until recently, involves a number of steps which we summarise here (leaving the reader to consult the excellent references, given at this end of this sub-section, for details). First, compute deviations of the series from a centered moving average. Second, break up the deviations into phases, according to the dates of cyclical peaks and troughs. Third, compute the mean values of the series for each successive phase, and smooth using three-item or two-item moving averages. The PAT is then given

by connecting these midpoints of these triplets or doublets; see Boschan and Ebanks (1978), Zarnowitz and Ozyildirim (2006) and Nilsson and Gyomai (2008) for details.

9.4.3 BN trends as limiting trends

As discussed in Section 9.4.1 above, the most general definition of Beveridge-Nelson trends is as long-run forecasts, absent deterministic growth, as the forecast horizon goes to infinity. The idea of using long-run forecasts to detrend a time-series is intuitive.

However, all possible trend components (i.e. trend components identified and estimated using any other decomposition method, including those reviewed above) in fact resemble (technically converge in expectation to) the BN trend as the forecast horizon increases.

This is seen by following Garratt et al. (2006) and considering any generic partitioning of y_t into trend μ_t^* and cyclical c_t^* components.

For any cyclical component it must follow that, in the long-run, the cyclical component is expected to be zero (otherwise it is not a cycle):

$$\lim_{h \to \infty} E(c_{t+h}^* | \Omega_t) = 0$$

Implying, given we have defined the vector of BN trends as $\mu_t = \mu_{t-1} + h + c(1)e_t$, that the long run forecast of the generic trend component μ_t^* equals μ_t:

$$\lim_{h \to \infty} E(\mu_{t+h}^* | \Omega_t) = \mu_t$$

Therefore all possible permanent components converge in expectation on the BN trends as the forecast horizon, h, increases. Alternative multivariate techniques, which we now discuss, that introduce additional assumptions (such as imposing long run restrictions) essentially redistribute some additional stationary element between the BN trend and cyclical components.

9.5 Univariate (or non-structural) versus multivariate (more structural) methods

9.5.1 Theoretical foundations

Multivariate estimators give a more economic interpretation to the cycle. Most commonly the cycle of interest is the output gap - i.e. the cyclical component of real GDP. Multivariate detrending methods essentially combine the estimators of the growth cycle with additional economic information. The Phillips Curve, for example, suggests that inflation data contain information about the output gap, while Okun's Law suggests unemployment is important. These economic variables may contain useful information about the supply side of the economy and the stage of the business cycle. The view is that output (or GDP) should not be detrended using output data alone.

Indeed, the output gap is larger the greater the degree of long-run forecastability; only when output growth is unforecastable does output follow a random walk. When output growth, to some extent, is forecastable there is a temporary component, i.e. there is an output gap, which in due course reverts to its unconditional mean (of zero). It is well known that multivariate models can, but need not always, deliver better forecasts of output

growth than univariate models. Therefore, what variables we include in the multivariate model, along with output, are of central importance in establishing how forecastable output growth is in the long run. In turn, this means that the variables we include in the multivariate model affect the behaviour of the cyclical series. It is not always clear what the relevant additional series to include are; use of multivariate methods is therefore not as mechanical as some univariate methods and does require the user to employ their judgement. Again this can be seen as a pro or a con.

Seven multivariate detrending models are reviewed:

   (i) multivariate unobserved components models (UC);

  (ii) multivariate Hodrick-Prescott models;

 (iii) multivariate nonparametric filters;

 (iv) multivariate Beveridge-Nelson decompositions;

  (v) structural VAR decompositions;

 (vi) production function approaches and

(vii) explicitly structural (DSGE) decompositions.

We provide only a cursory review of those multivariate methods that more closely resemble the analogous univariate filters reviewed above.

9.5.2 Multivariate UC models

The UC model seen above in section 9.4.1 can be extended to the multivariate case by considering the components of $y_t = \mu_t + c_t$ as N-dimensional vectors, with the co-variances between the disturbances to the unobserved components represented by $N \times N$ matrices. Carvalho et al. (2007) provide a recent general discussion of the use of multivariate UC models for business cycle analysis.

In the multivariate case a reduced number of shocks might drive the N trend or cyclical components. This implies restrictions on the rank of the covariance matrices in the multivariate UC model. As discussed for univariate UC models, the disturbances to the trend and cyclical components need not be uncorrelated.

In practice, popular multivariate UC models for the estimation of the output gap are either bivariate (GDP and inflation) or trivariate (GDP, inflation and unemployment); see Gerlach and Smets (1999) and Gordon (1997). Typically some care needs to be taken in defining the model to ensure the cyclical estimates look reasonable. To avoid having to fix the signal to noise ratio, assumptions are often made about the nature of the cyclical processes. Otherwise, the trend component can account for all the variation in the level of the variable and soak up all residual variation. Estimation of these models is more complicated, but again facilitated by software like STAMP.

Multivariate UC models which give a nonparametric interpretation to the cycle are also used to detrend. The model of Valle e Azevedo (2006), which also allows for phase shifts in the cyclical components of multiple time series, is a model-based multivariate band pass filter. Creal et al. (2010) allow for time-variation and stochastic volatility which is critical when modelling many macroeconomic samples.

Multivariate correlated UC models (e.g. see Basistha and Nelson (2007)) relax the assumption that trend output follows a random walk by using economic theory (specifically the New Keynesian Phillips Curve) to identify trend output as consisting of both a random walk component and a stationary component.

9.5.3 Multivariate HP trends

Laxton and Tatlow (1992) proposed an extension to the Hodrick-Prescott (HP) filter which incorporates economic information. Additional, so-called economic, constraints are imposed on the minimisation, see equation (9.5) from which the HP filter is defined. The residuals from a structural equation, such as the Phillips Curve or Okun's Law, are added to the minimisation problem that the univariate HP filter seeks to solve. But just as the univariate or traditional HP filter can be interpreted within an unobserved components framework (see Harvey and Jaeger (1993)), so can the multivariate HP filter. This facilitates estimation by maximum likelihood and inference since confidence bands around the estimates can be derived from the Kalman filter recursions. Harvey (1989) provides details of these recursions. STAMP or the underlying matrix programming language (Ox) provide one means of estimating multivariate HP trends. Use of these software is not straightforward for non-experts.

9.5.4 Multivariate nonparametric filters

In turn, Valle e Azevedo (2011) extends Christiano and Fitzgerald (1999) to the multivariate case. Valle e Azevedo (2011) develops an approximation to the ideal business cycle that incorporates information from an arbitrary number of time series and explores how additional information can reduce the uncertainty associated with real-time estimates of the business cycle. The filter is an optimal (in the mean squared error sense) approximation to the ideal filter and isolates a specified range of fluctuations in a time series, e.g., business cycle fluctuations in macroeconomic time series. But the method does require knowledge of the true second-order moments of the data.

Wildi (2008) has a similar aim but does not make any assumptions about the specific data-generating process as the author generalises univariate to multivariate band pass filters, so a larger set of relevant data can be used to help detrend. Importantly, cointegration (long-run level) relationships can be imposed between the series of interest.

Application and estimation of multivariate nonparametric filters is more complicated than their univariate counterparts. It requires specialist software that is not typically publicly available or built into commercial software.

9.5.5 Multivariate BN trends

Multivariate Beveridge-Nelson (BN) trends are usually derived from the vector moving average (MA) representation of a VAR model. Proietti (1997) and Garratt et al. (2006) show that, alternatively, they can be derived directly from the VAR representation. This lets them relate the cycle directly to the underlying stationary processes, which have an economic interpretation, which drive the cycle.

Importantly, the BN trend component need not be smooth as typically implied by statistical decomposition methods, like UC models. The explicit use of economic theory when decomposing a time-series affects the properties of the trend and cyclical components. Consider a generic VAR model in an m-vector of typically nonstationary time-series, y_t:

$$A(L)y_t = \mu_t$$

ignoring, for convenience only, deterministic terms. This VAR can be written as an infinite order vector Moving Average (MA) process in the first differences, Δy_t

$$\Delta y_t = C(L)\mu_t$$

The traditional multivariate Beveridge and Nelson decomposition of y_t is then defined by decomposing the matrix lag polynomial $C(L)$ into its trend and cyclical components:

$$C(L) = C(1) + (1 - L)C^*(L)$$

In practice, many economic time-series move together in the long-run. They are said to be cointegrated. Cointegration of order r is defined when the matrix of long run multipliers is of reduced rank:

$$rank[C(1)] = m - r.$$

Imposing this reduced rank restriction on the vector MA, and solving for the level of y_t yields the common trends representation or the multivariate BN decomposition:

$$y_t = C(1) \sum_{j=1}^{t} \mu_j + C^*(L)\mu_t$$

where $C(1)\Sigma_{j=1}^{t}\mu_j$ represents the trend or long run component of y_t; the shocks that drive the common stochastic trends are permanent.

9.5.6 Structural VAR (SVAR) decompositions

Following Blanchard and Quah (1989), VAR models with economic restrictions imposed on the long-run have been used to estimate output gaps. In the SVAR approach identifying restrictions are imposed on the matrix of long-run multipliers so that demand (or transitory) shocks can be distinguished from supply (or permanent shocks). The output gap is the cumulative sum of the transitory shocks to output. Potential output is then the cumulative sum of the permanent shocks. Importantly this means that, in contrast to other multivariate decomposition methods such as Beveridge-Nelson trends and traditional multivariate Unobserved Components models, potential output is not assumed to follow a random walk. Potential output has a random walk component but also an additional stationary component, reflecting the short-run dynamics of the permanent shocks.

The structural VAR relates the (reduced form) VAR errors μ_t to structural errors, ε_t, deemed to have some economic interpretation by assuming $\varepsilon_t = A_0\mu_t$. In a stylised economic model, we might classify the structural shocks in ε_t into two groups: the "demand" shocks, which are transitory, and the "supply" shocks, which are permanent. The trend component is then defined with respect not to all the shocks, nor with respect to μ_t as in the BN decomposition, but only the "supply" shocks. We can see this more formally by considering the structural vector MA process

$$y_t = C(L)A_0^{-1}\varepsilon_t = D(L)\varepsilon_t.$$

Structural VAR based detrending methods typically then identify "structural" shocks by restricting the matrix of structural long run multipliers, $D(L)$. These restrictions conventionally take the form of assuming $D(1)$ is lower triangular. Motivated by "long run" *a priori* theory (e.g. a long run neutrality hypothesis) a shock is attributed economic meaning on the basis of whether its long run (infinite horizon) impact on particular endogenous variables is permanent or transitory.

SVAR based trend-cycle decompositions then proceed as with the multivariate BN decomposition but work off the structural shocks ε_t not the reduced form shocks μ_t

$$y_t = D(1) \sum_{j=1}^{t} \varepsilon_j + D^*(L)\varepsilon_t$$

The SVAR then defines the trend component as the cumulative sum of the permanent shocks in ε_t.

What the SVAR model has done is add in some additional stationary component to the BN trend by imposing additional identification restrictions; see also Dupasquier et al. (1999) for a survey. This is seen in the equation above by simply noting that in the SVAR method we define the trend component as the sum of the permanent shocks. Having identified the structural shocks we can see that the SVAR trend will involve summing elements of both $D(1)\Sigma_{j=1}^{t}\varepsilon_j$ and $D^*(L)\varepsilon_t$, since the latter picks up the (short-run) dynamics of the permanent shocks. That is, the permanent shocks fall both in $D(1)\sum_{j=1}^{t}\varepsilon_j$ and $D^*(L)\varepsilon_t$. Hence in SVAR models the trend is not restricted to be a random walk, although it has a random walk component.

As with the multivariate BN decomposition, considered above, cointegrating restrictions can also be imposed on the underlying VAR model when present. In this case a reduced number of shocks drive the common stochastic trends. In fact, this leads to a sub-class of detrending methods.

Structural VAR decompositions with cointegration

Rather than detrend a vector of time-series having estimated a stationary (typically first-differenced) VAR model a cointegrating VAR model is also commonly considered. Cointegrating VAR models have the perceived attraction that their long-run behaviour is consistent with economic theory.

A popular cointegrating VAR model is based on King et al. (1991) and in its simplest setting considers a three variable system of GDP, consumption and investment. There are believed to be two cointegrating vectors: the ratios of consumption and investment to GDP are deemed stationary. Therefore, there is one permanent shock and two transitory shocks. The cyclical component of GDP, the "output gap", is the cumulative sum of the two transitory shocks to output. DSGE restrictions, discussed further below, are also imposed on the structural VAR model parameters as long-run restrictions; e.g. see Pesaran and Smith (2006).

In practice, the estimates of trend from SVAR models need not be smooth as typically implied by more statistical decomposition methods, like UC models. The explicit use of economic theory when decomposing a time-series affects the properties of the trend and cyclical components. Software to compute SVAR based estimates of the cycle is not as freely available as for some of the other detrending methods reviewed in this Chapter; but matrix programming languages like Gauss or Matlab facilitate estimation and programmes can be found on the web. But the reliability and certainly ease-of-use of these is arguably less than for simpler detrending methods and/or when a commercially available software package is available (such as EViews or STAMP).

The importance of the lag length in the VAR

The long-run restrictions implicit to identification require estimation of the matrix of long run responses (the sum of the lag polynomial in the moving-average representation corresponding to the VAR model). This is based on estimation of the sum of the coefficients in the VAR model. Therefore the reliability of SVAR estimates of the output gap rests on reliable estimation of the AR coefficients and this also depends on the lag order chosen; see Faust and Leeper (1997) and DeSerres and Guay (1995).

Of course, a VAR model can always be written in state-space form. So, in this sense, there is no distinction between VAR and state-space models. Proietti (1997) derives expressions for the unobserved components of the multivariate BN decomposition from a SVAR model. This involves writing the cointegrating VAR model in state-space form, and then deriving expressions for the trend and cycle based on the Kalman filter.

We have seen in section 9.4.3 that while the random walk feature of BN trends is often argued to be a disadvantage of BN trends, in fact all possible trend components (i.e. trend components identified and estimated using any multivariate decomposition method including SVAR methods) in fact converge in expectation to the

BN trend as the forecast horizon increases. Only in the short-run, therefore, does the SVAR approach ensure the trend no longer follows a random walk. This follows from the property that the SVAR essentially adds to the random walk component of the trend some of the stationary movements in the permanent component.

9.5.7 Production function approaches

Economists find it useful to describe the supply side of the economy using the concept of the production function, which relates inputs of capital and labour to the level of output. A production function based measure of the output gap, such as that discussed in Cotis (2003), compares actual inputs of labour and capital to potential equilibrium levels, and calculates a gap between them on this basis. Although academically appealing, this approach is fraught with problems. It is very difficult to assess the equilibrium level of labour supply and unemployment. The most common way to model the evolving labour market equilibrium, or the NAIRU, is to utilise time dependent intercepts reflecting the changing nature of these markets. Production function based methods of measuring output gaps commonly model these evolving intercepts as unobserved components and extract them via the Kalman filter, for instance. Considerable judgement is required in specifying the appropriate unobserved components model that is to be used in the construction of the production function measure of equilibrium output. Proietti et al. (2007) provide a model-based interpretation of the production function approach.

9.5.8 DSGE Models and the Output Gap

Another class of model increasingly used by economists to understand business cycle movements is Dynamic Stochastic General Equilibrium (DSGE) models. DSGE models impose a greater degree of economic content on the notion of trend output, and in turn on the output gap. In contrast, UC models and SVAR models impose less economic theory restrictions and allow the data to determine, to a greater extent, the nature of trend output. As a result, these more statistical filters might be suspected to suffer less from misspecification than DSGE models. Typically, SVAR models impose more economic theory restrictions on trend output than UC models, and distinguish between long-run and short-run restrictions, so in this sense can be seen to lie between DSGE and UC models in terms of their economic content.

The use of estimated DSGE models for business cycle analysis offers a number of putative advantages. Principally, since the DSGE model has a micro-motivated economic structure it allows one to understand, from an economic theory perspective, macroeconomic deviations from trend, i.e. business cycle dynamics and also model the transmission of shocks. For example, one can use a DSGE model to decompose business cycle fluctuations into different shocks, which have an economic interpretation, and thereby understand the causes of business cycle fluctuations. The DSGE model facilitates estimation of the structural shocks and of the parameters deemed to explain economic behaviour. Estimation is facilitated by software like Dynare.

But DSGE models embed different notions of trend, depending on the restrictions imposed on the underlying statistical model, a VARMA model. The different trend concepts that might be differentiated include efficient output, potential output, the natural rate of output and long-run output; see Edge et al. (2008), Justiniano and Primiceri (2008) and Vetlov et al. (2011). This contrasts most statistical detrending methods which imply a unique output gap estimate.

Many DSGE models make strong and apparently economically, rather than statistically, motivated assumptions about the nature of the trends in the economic variables in the DSGE model. For example, in some DSGE models, time-series are Hodrick-Prescott detrended off-model, prior to estimation of the DSGE model. The Smets and Wouters (2007) DSGE model assumes a balanced growth path whereby the trends in the real variables in the model are driven by deterministic labour augmenting technology progress. This implies a single common (deterministic) trend drives income, consumption and investment. This assumption is at odds with much statistical evidence (at least from a frequentist perspective). More recently attempts have been

made to robustify inference in the DSGE model to treatment of the trend components; e.g. see Cayen et al. (2009) and Canova and Ferroni (2011).

In practice, cycles from DSGE models can have quite different properties from many more statistical (or at best mixed) approaches to estimation of the trend; e.g. see Neiss and Nelson (2005) and Edge et al. (2008). This is because in DSGE models trend output might be affected by real shocks over the business cycle. This implies that it does not follow a smooth trend, as in many statistical decompositions such as UC models. Government spending, terms of trade shocks, tax changes and taste shocks all affect households' decisions about consumption and so affect labour supply and hence trend output. Trend output, therefore, can vary over the business cycle.

9.6 Linear versus nonlinear detrending models

Time-series can also be detrended using models which allow the series of interest to be nonlinear rather than linear. A popular nonlinear model, the Hamilton (1989) model, is to let the growth rate of the series of interest have a Markov switching mean. Although the cycle is not identified, when the Hamilton model is written in UC form, one can still estimate the trend component (and thereby the growth cycle) for the Markov Switching model using a generalisation of the Beveridge-Nelson (BN) decomposition suggested by Morley and Piger (2008). Importantly the decomposition of Morley and Piger (2008) allows for the fact that the trend component need not have a constant average growth rate as implied by the traditional BN decomposition even when applied to Markov Switching models. One can also detrend using identified nonlinear UC models, such as the model of Lam (1990) which allows for regime shifting in the drift parameter.

An alternative approach is to postulate that the regime switching shows up in the cyclical component itself. This delivers Friedman's "plucking model" proposed by Kim and Nelson (1999). Friedman's plucking model of business fluctuations suggests that output cannot exceed an upper limit, but it is occasionally "plucked" downward below trend as a result of economic recessions. Kim et al. (2005) capture strong output growth in the aftermath of recessions by exogenously allowing growth dynamics to change in the quarters immediately after a decline in output below its historical maximum. Thereby they extend the Hamilton model to allow for a post-recession bounceback. An alternative approach is to allow for a third regime. Sinclair (2009) generalises Kim and Nelson (1999) to consider non-zero correlations between the permanent and transitory innovations in a plucking model.

9.7 Accommodating uncertainty when detrending

As seen there exist many alternative econometric methods to identify and estimate trends and cycles. There is, effectively, "model uncertainty". This manifests itself at a practical level in different, and potentially contrasting, quantitative assessments about the trend and cycle. Since inference about trends and cycle can be sensitive to measurement, it is important to compare and contrast estimates obtained using alternative methods.

This provides a rationale for model-averaging. Garratt et al. (2014) suggest that when the output gap is to be used for a specific purpose, like forecasting inflation, one should combine alternative estimators according to their ability to forecast inflation. This should be done acknowledging the uncertainty in individual cyclical estimates by considering their densities, not just point estimates of the cycle. This is because there is "within" as well as "between" model uncertainty. As Mitchell (2007) showed, even for a single detrending method, uncertainty bands around point estimates of the output gap can be very wide, when computed in real-time. Morley and Piger (2008), within a Bayesian approach, also combine alternative output gap estimates giving a higher weight to those gap estimates from better (in-sample) fitting models.

9.8 Performance in real-time: The "end point" problem

9.8.1 The nature of real-time or end-of-sample cyclical estimates

Real-time or end-of-sample cyclical estimates can be unreliable, in the sense that there is often in practice found to be a large and significant revision or forecasting error; see Orphanides and van Norden (2002) for an application to the US output gap, and Mitchell (2007) and Marcellino and Musso (2011) to the Euro-Area (EA). But policymakers require cyclical estimates in real-time. They do not have the luxury of being able to wait before deciding whether the economy is currently lying above or below its trend level. Their problem can be interpreted as a forecasting one, since these real-time cyclical estimates are forecasts, in the sense that they are expectations of the cycle conditional on incomplete information. Only with the arrival of additional information, such as revised historical data and data not available at the time, do the cyclical estimates eventually settle down at their "final" values.

Revisions to end-of-sample cyclical estimates are explained by both "data" and "statistical" revisions. In general, data revisions, explained by revisions to published data, have been found to be less important than so-called statistical revisions. Statistical revisions are explained by the arrival of new data helping the modeller, with the advantage of hindsight, better understand the position of the cycle.

In the absence of data revisions, revisions to real-time output gap estimates are explained solely by forecasting errors. As discussed, detrending filters, whether interpreted as parametric or nonparametric, can be seen to involve application of a moving-average filter to the raw data. Most detrending filters, such as the HP filter, imply a smooth two-sided moving average in the middle of the sample with no time-series observation receiving a large weight relative to its close neighbours. However, at the end of the sample the moving average becomes firstly asymmetric and then one-sided. It is well known that application of a one-sided filter will lead to more volatile estimates at the end of the sample as the weight on the last observation will be much higher than any of the weights associated with application of the filter in the middle of the sample since at the end of the sample the filter is unable to distinguish temporary from permanent shocks. To apply the two-sided filter future values need to be forecasted. It follows that if these forecast errors are reduced, the revision will decrease. Forecasting future values prior to detrending facilitates use of a less one-sided de-trending filter.

For a correctly specified UC model, application of the one-sided Kalman filter is, in fact, equivalent to application of the two-sided filter (smoother) to y_t extended infinitely into the future with optimal forecasts. These optimal, minimum mean square error (MSE), forecasts are derived via the Kalman filter, exploiting the state-space representation for a given output gap estimator. There is therefore no need for forecast extensions.

Only when the UC model is misspecified can forecast extensions, using an alternative forecasting model, help to attenuate revisions (lessen estimation MSE). The importance of real-time testing for model misspecification is therefore apparent.

9.8.2 Improving the reliability of end-of-sample cyclical estimates

Growing attention has therefore been devoted to understanding revisions (e.g. see Proietti (2009)) and improving the reliability of end-of-sample cyclical estimates through forecasting and modelling revisions. Garratt et al. (2008), for example, employ cointegrating VAR techniques to model and forecast real-time data prior to detrending. Clements and Galvão (2013) use vector-autoregressive models of data vintages to provide forecasts of post-revision values of future observations and thereby improve real-time cyclical estimates. Others, using some of the multivariate detrending methods discussed above, consider whether multivariate models deliver real-time output gap estimates with fewer revisions; e.g. see Planas and Rossi (2004).

9.9 Evaluating alternative detrending methods

The fundamental problem associated with the putative evaluation of alternative trend and cyclical estimates is that the true "trend" and "cycle" are never observed, even *ex post*. While economic theory is sometimes believed to offer guidance about the "true" processes for trend and cycle, in practice there is also uncertainty about the "right" economic model. This means that usually there is no consensus about the "outturns" which more statistically based detrending methods are meant to proxy.

Therefore, in practice, two main evaluation criteria are used. First, revisions to real-time cyclical estimates are minimised. But we should stress that large revisions need not indicate poor quality. Secondly, the quality of cyclical estimates is evaluated according to their purpose. For example, output gap estimates are often used to assess inflationary pressures. If so, the quality of alternative output gap estimates can be evaluated according to their ability to forecast inflation. Or they can be evaluated according to their ability to forecast future output growth. As argued by Nelson (2008), when the output gap is positive (output is above trend) we are requiring future growth at a below trend rate. Analogously, if the output gap is negative recovery requires future growth at an above trend rate. As Nelson (2008) says, "predictability is the essence of transitory variation".

Finally, we can stress that a comparative analysis among various detrending techniques makes sense only if we are considering techniques laying in the same family, for example comparing deterministic and stochastic detrending techniques does not help too much in assessing thier quality because the underlying definition of "cycle" is very different. Furthermore, detrending techniques aiming at extracing the cyclical component (band pass filters) produce smoother results than those techniques aiming to estimate first the trend (low pass filters). This is due to the fact that in the first category of detrending methods the detrending method is associated to the trend while in the second group this is associated to the cyclical component which by the way appears less smooth.

9.10 The aggregate and disaggregate European growth cycle over the recent recession: A real-time illustration

Theoretical foundations

Finally, to illustrate the use of detrending methods in practice we detrend Euro-area GDP. Our principal objective is to demonstrate: (i) that inference, about the cycle, is sensitive to the chosen detrending method; (ii) that real-time estimates of the cycle can be extremely unreliable in the sense that they can be subject to considerable revision; and (iii), that it is prudent to consult various measures of the cycle, since different evaluation criteria provide alternative rankings of different approaches to estimation of the cycle. In other words, we illustrate that there is no single detrending measure that consistently dominates across alternative evaluation criteria. Indeed, the ranking of different detrending methods is likely sample and variable specific. There is not a single 'best' detrending method which users with different evaluation criteria will agree upon.

In the course of the application we draw out another important issue, namely detrending with aggregated data.

Since Euro-area GDP involves the aggregation of GDP data for each of the member countries, one can estimate the Euro-area growth cycle using three different approaches. First, the "direct" approach, which consists of decomposing the raw data of the aggregate itself into trend and cyclical components. Secondly, the "indirect" approach, which consists of decomposing the raw data corresponding to the sub-components (national GDP) and then aggregating the county-level cycles using their current quarter shares of aggregate GDP as

weights. Thirdly, the "multivariate" or simultaneous approach, which involves decomposing the disaggregates simultaneously into trend and cyclical components. In general, these approaches will deliver different aggregate estimates of the (growth) cycle.

We consider the real-time behaviour and performance of all three of these decomposition methods, using real-time vintage data for both the Euro-area and the individual countries. We focus on the use of the Beveridge-Nelson decomposition method as a unified means to extract growth cycle estimates from (V)ARMA models. The BN decomposition conveniently allows one to extract cyclical estimates from both univariate and multivariate models; what differs is the conditioning information. This is important, as while we do not compare all of the many aforementioned detrending methods, comparing and constrasting cyclical estimates according to the information set deployed is sufficient to illustrate that inference about the cycle is indeed extremely sensitive to how the cycle is measured. Following Mitchell, Robertson and Wright (2014), we stress why ARMA models should be used in particular in univariate decompositions, since they actually bridge the gap relative to multivariate filters given that they can be seen to accommodate the potential dependence of any observed or unobserved variable or variables on the variable of interest. In other words, they can be seen as the univariate reduced form underlying any generic multivariate decomposition method.

Application: Detrending Euro-area GDP

We undertake out-of-sample simulations using real-time (vintage) data to compare the revisions properties of direct, indirect and multivariate approaches to estimating the cyclical component of GDP in the Euro-area (defined as the largest nine EA countries where consistent real-time data are available: Austria, Belgium, Germany, Spain, Finland, France, Greece, Italy and The Netherlands). This involves for the direct and indirect approaches recursively estimating univariate ARMA models in GDP growth. Given the lag order of the underlying ARMA model in GDP growth is known to affect inference we consider two variants. First we consider AR(8) models as in Garratt et al. (2008). Secondly, we consider ARMA(1,1) models as suggested by Mitchell, Robertson and Wright (2014). The multivariate estimates are the multivariate BN cycles from a 9-variable VAR (of lag order 2) in the 9 European countries' GDP with and without common trend restrictions imposed. These common trend restrictions allow for the possibility that there might be common European cycles driving national cycles, which would induce synchronous behaviour between the nine countries. Given there is uncertainty about the specific number of common trends, for robustness we consider ten different VAR models with cointegrating rank values $r = 0, \ldots, r = 9$. To provide an indication of what the cyclical estimates for Euro-area GDP look like from these three approaches, Figure 9.1 plots the direct, indirect and multivariate estimates of the EA cycle using the 'final' vintage (2010q4) data.

Figure 9.1 shows that while the direct and indirect approaches give a similar picture about the European cycle, with correlation coefficients at 0.83 or higher, the multivariate approach can paint quite a different picture. But this does depend on the number of cointegrating vectors imposed in the VAR.

The correlation between the VAR-based cycle and the direct and indirect estimates declines as the number of cointegrating vectors rises above 3 or 4. That is, imposing more and more commonality on the VAR reduces the similarity with the direct and indirect cycles. But there is evidence that imposing a degree of commonality does increase correlation against the direct and indirect estimates, since a cointegrating rank of about 3 tends to deliver increased correlation relative to the case of assuming less or no cointegration.

We then undertake out-of-sample simulations, across 40 out-of-sample data vintages, to evaluate the revisions properties and performance of real-time EA9 output gap estimates over the period 2000q4 to 2010q3.

To provide a graphical indication of the estimates Figure 9.2 plots the real-time estimates alongside the final ones. It is seen that analysis is dominated by the 2007-8 recession. There is a general tendency for both the direct and indirect methods to overstate the size of the (negative) output gap in the recession. Typically they indicated that output was 5% or more below trend in real-time; when the final vintage estimates suggest a more moderate deterioration of around 2%. This is because, with the advantage of hindsight, much of the

Figure 9.1: Direct, Indirect and Multivariate based estimates of the Euro-area (9 countries) Beveridge-Nelson cycle using final vintage data
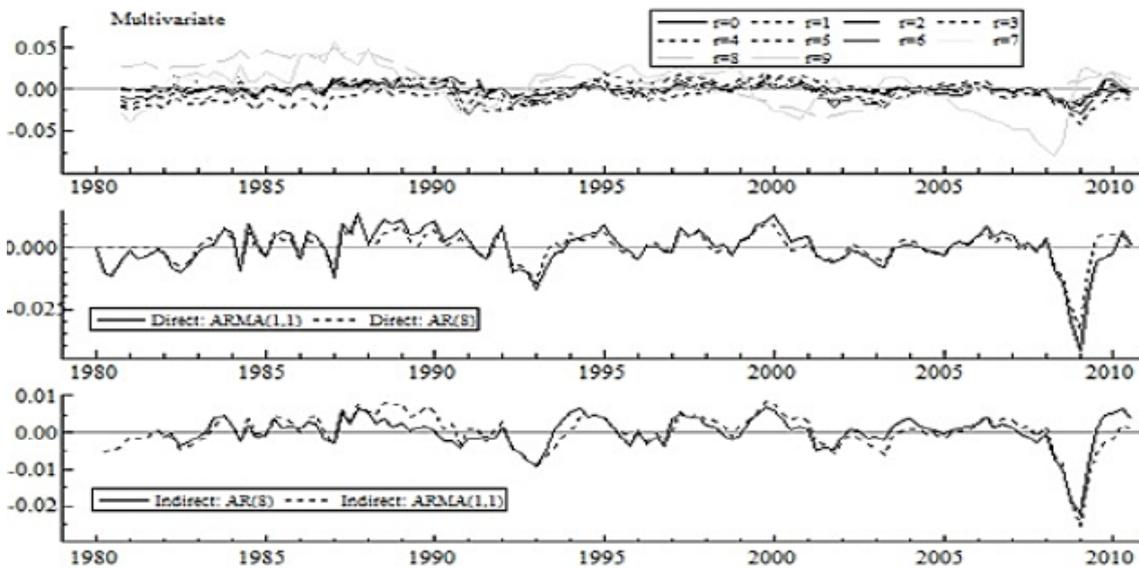
dramatic fall in European GDP is attributed to a trend movement, rather than a cyclical movement. It takes some time after the event for the trend estimates to catch up with the actual movements in GDP. However, the multivariate approach does not suffer from this problem, although revisions are in fact often larger as there are more persistent differences between the real-time and final estimates than for the direct and indirect approaches.

Figure 9.2: Real-time and final vintage (marked) output gap estimates (for the EA9) using direct, indirect and (selected) VAR Beveridge-Nelson detrending methods
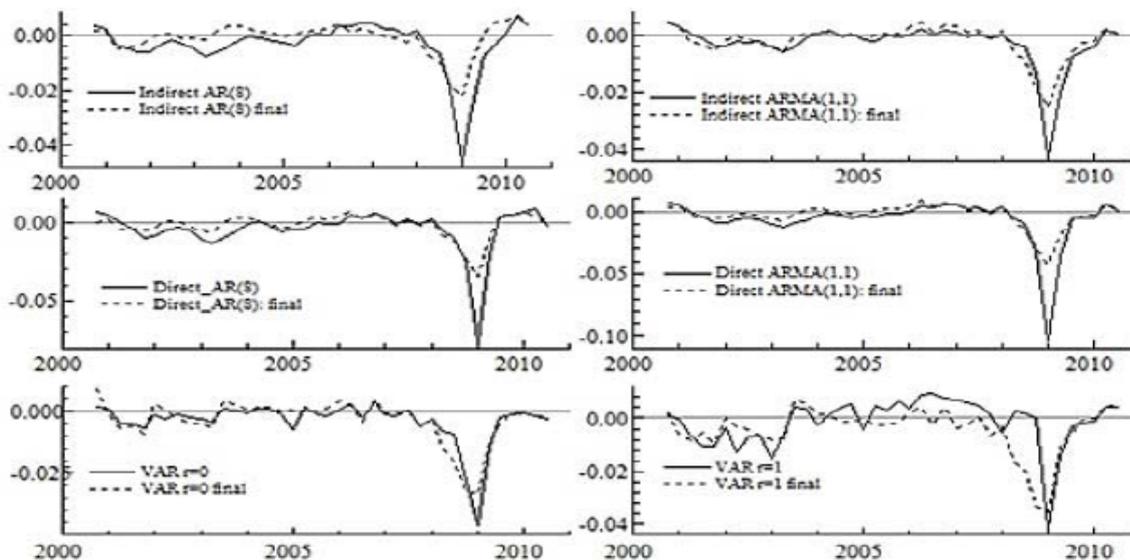
Table 9.1 summarises the results further by reporting some evaluation statistics. It evaluates firstly the revisions behaviour of the real-time estimates, by comparing the real-time estimates with the final ones (defined as 2010q4 vintage data). Evaluation involves looking to see if there is evidence for bias in the revisions. This involves undertaking a t-test using a Newey-West robust estimate of the standard error. We also report the RMSE of the revision; and the RMSE relative to the indirect approach (using an ARMA(1,1) process). Secondly, we report the correlation of both the real-time and final vintage output gap estimates against HICP

Table 9.1: Real-time revisions to EA output gaps. Revisions behaviour and correlation of the real-time and final vintage output gaps against inflation for the indirect, direct and multivariate VAR models (for differing cointegrating ranks, r)

		Mean	Bias	RMSE	Rel. RMSE	Corr vs. infl.	Corr vs. infl.
		×100	tstat			realtime	final
Indirect	AR(8)	0.25	1.95	0.037	1.65	0.17	-0.03
Indirect	ARMA(1,1)	0.1	1.66	0.022	1	0.29	0.21
Direct	AR(8)	0.27	1.95	0.05	2.22	0.08	0.04
Direct	ARMA(1,1)	0.37	2.17	0.068	3.01	0.2	0.2
VAR	r=0	0.03	0.62	0.024	1.05	0.18	0.13
VAR	r=1	-0.28	-1.64	0.054	2.4	-0.01	0.06
VAR	r=2	-0.17	-0.56	0.074	3.27	0.12	0.39
VAR	r=3	-0.77	-1.62	0.123	5.48	-0.11	0.53
VAR	r=4	-0.47	-0.48	0.221	9.86	0.06	0.62
VAR	r=5	0.64	0.49	0.281	12.53	0.15	0.52
VAR	r=6	-0.23	-0.23	0.224	9.97	0.12	0.3
VAR	r=7	-2.24	-2.66	0.23	10.25	-0.02	0.23
VAR	r=8	-2.49	-5.44	0.19	8.48	0.48	0.35
VAR	r=9	-1.72	-3.73	0.155	6.89	0.47	0.34

inflation, one-quarter ahead.

Inspection of Table 9.1 shows that the indirect method delivers real-time estimates with fewer revisions than the direct method. The more parsimonious ARMA(1,1) model delivers estimates with lower RMSE than the AR(8) model. The indirect approach also delivers estimates with a lower RMSE than the majority of the multivariate (VAR-based) decomposition methods. As common trends are imposed and $r > 1$, the revisions behaviour of the real-time VAR-based estimates tends to deteriorate.

However, demonstrating once again the dangers of evaluating real-time cyclical estimates according to their revisions properties only, Table 9.1 shows that the ability of the cyclical estimates to explain inflation one-quarter ahead does not match their revisions performance. The direct and indirect approaches, both in real-time and at the end of the sample, exhibit at best a weak correlation with inflation. However, some of the multivariate estimates do deliver estimates with higher explanatory power for inflation, despite their larger revisions.

9.11 Conclusions

This Chapter reviews the large detrending literature. It classifies different detrending methods; and discusses their relative advantages. This is important, as inference is sensitive to measurement, so users of trend and cyclical estimates need to be aware of different methods' pros and cons and that there is no clear consensus as to the preferred (single) detrending method. Indeed, since the true "trend" and "cycle" are never observed, even *ex post*, this Chapter argues that it is not clearcut how one should evaluate alternative empirical estimates of the trend and cycle. In an illustrative application we remind readers that real-time estimates of the cycle can be extremely unreliable, in the sense that they can be subject to considerable revision. We also suggest that it is prudent to consult various measures of the cycle, since different evaluation criteria provide alternative rankings of different approaches to estimation of the cycle.

Bibliography

Basistha A., Nelson C.R. (2007), New Measures of the Output gap Based on the Forward-Looking New Keynesian Phillips Curve, Journal of Monetary Economics, 54, 498-511.

Baxter M., King R. G. (1999), Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series, Review of Economics and Statistics, 81(4), 575-593.

Beveridge S., Nelson C.R. (1981), A New Approach to Decomposition of Time Series into Permanent and Transitory Components with Particular Attention to Measurements of the Business Cycle, Journal of Monetary Economics, 7, 151-174.

Blanchard O.J., Quah D. (1989), The dynamic effects of aggregate demand and supply disturbances, American Economic Review, 79, 655-73.

Boschan C., Ebanks W.W. (1978), The Phase-Average Trend: A New Way of Measuring Growth, in 1978 Proceedings of the Business and Economic Statistics Section, American Statistical Association, Washington, D.C.

Bry G., Boschan C. (1971), Cyclical Analysis of Time Series: Procedures and Computer Programs, New York, NBER.

Canova F. (1998), Detrending and Business Cycle Facts, Journal of Monetary Economics, 41, 475-512.

Canova F., Ferroni F. (2011), Multiple Filtering Devices for the Estimation of Cyclical DSGE Models Quantitative Economics, 2 (1), 2011, 73-98.

Carvalho V., Harvey A., Trimbur, T. (2007), A note on common cycles, common trends, and convergence, Journal of Business & Economic Statistics 25, 12-20.

Cayen J-P., Gosselin M-A., Kozicki S. (2009), Estimating DSGE-Model-Consistent Trends for Use in Forecasting, Working Papers 09-35, Bank of Canada.

Chagny O., Döpke J. (2001), Measures of the Output Gap in the Euro-Zone: An Empirical Assessment of Selected Methods, Kiel Working Papers 1053.

Clements M.P., Galvão A. (2013), Improving Real-time Estimates of Output Gaps and Inflation Trends with Multiple-vintage VAR Models, Journal of Business and Economic Statistics, 30, 554-562.

Christiano L., Fitzgerald T.J. (2003), The Band Pass Filter, International Economic Review, 44, 2, 435-465.

Cogley T. (2001), Alternative definitions of the business cycle and their implications for business cycle models: A reply to Torben Mark Pederson, Journal of Economic Dynamics and Control, 25(8), 1103-1107.

Cogley T., Nason, J.M. (1995), Effects of the Hodrick-Prescott filter on trend and difference stationary time series: implications for business cycle research, Journal of Economic Dynamics and Control, 19, 1-2, pp. 253-78.

Cotis J-P. (2003), Methods of Measuring Potential Output: Benefits and Pitfalls from a Policy Perspective, OECD working paper.

Creal D., Koopman S-J., Zivot (2010), Extracting a robust US business cycle using a time-varying multivariate model-based band-pass filter. Journal of Applied Econometrics, 25, 695-719.

DeSerres A., Guay, A. (1995), Selection of the truncation lag in structural VARs (or VECMs) with long run restrictions, Bank of Canada Working Paper 95-9.

Dupasquier C., Guay A., St-Amant P. (1999), A survey of alternative methodologies for estimating potential output and the output gap, Journal of Macroeconomics, 21, 577-595.

Edge R.M., Kiley M.T., Laforte J.P (2008), Natural Rate Measures in an Estimated DSGE Model of the US Economy, Journal of Economic Dynamics and Control, 32, 2512-35.

Faust J., Leeper E.M. (1997), When do long run identifying restrictions give reliable results? Journal of Business and Economic Statistics, 15, 345–353.

Garratt A., Lee K., Mise E., Shields K. (2008), Real-Time Representations of the Output Gap, Review of Economics and Statistics, 90(4), 792-804.

Garratt A., Mitchell J., Vahey S.P. (2014), Measuring Output Gap Nowcast Uncertainty, International Journal of Forecasting, 30, 268-279.

Garratt A., Robertson D., Wright S. (2006), Permanent vs. Transitory Components and Economic Fundamentals, Journal of Applied Econometrics, 21, 521-542.

Gerlach S., Smets F. (1999), Output gaps and monetary policy in the EMU area, European Economic Review, 43, 801-812.

Gordon R J. (1997), The time-varying NAIRU and its implications for economic policy, Journal of Economic Perspectives, 11, 11-32.

Hamilton J.D. (1989), A new approach to the analysis of nonstationary time series and the business cycle, Econometrica, 57, pp. 357-84.

Harding D., Pagan A. (2002), Dissecting the cycle: a methodological investigation, Journal of Monetary Economics, 49(2), 365- 381.

Harvey A.C. (1985), Trends and Cycles in Macroeconomic Time Series, Journal of Business and Economic Statistics, vol. 3, pp. 216-230.

Harvey A.C. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Harvey A.C., Jaeger A. (1993), Detrending, stylized facts, and the business cycle. Journal of Applied Econometrics 3, 231-247.

Harvey A.C., Koopman S.J. (2000), Signal extraction and the formulation of unobserved components models, Econometrics Journal, 3, 1, pp. 84-107.

Harvey A.C., Trimbur T. (2003), General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series, Review of Economics and Statistics, 85(2), 244-255.

Harvey A.C., Trimbur T.M., Van Dijk H.K. (2007), Trends and cycles in economic time series: A Bayesian approach, Journal of Econometrics, 140(2), 618-649.

Hodrick R.J., Prescott E.C. (1980), Postwar U.S. Business Cycles: An Empirical Investigation, Carnegie Melon University discussion paper no. 451.

Hodrick R.J., Prescott E.C. (1997), Post-War U.S. Business Cycles: An Empirical Investigation, Journal of Money, Banking and Credit, 29, 1-16.

Justiniano A., Primiceri G.E. (2008), Potential and Natural Output, Working Paper, Federal Reserve Bank of Chicago, June.

Kim C-J., Morley J., Piger J. (2005), Nonlinearity and the permanent effects of recessions, Journal of Applied Econometrics, 20, 2, 291-309.

Kim C-J., Nelson C.R. (1999), Friedman's plucking model of business fluctuations: Tests and estimates of permanent and transitory components, Journal of Money, Credit, and Banking, 31, 3, 317-34.

King R.G., Plosser C., Stock J., Watson M. (1991), Stochastic trends and economic fluctuations, American Economic Review, 81, 819-840.

King R.G., Rebelo S. (1993) Low Frequency Filtering and Real Business-Cycles Journal of Economic Dynamics and Control, 17, 207-31.

Lam P.S. (1990), The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series, Journal of Monetary Economics, 26, 3, pp. 409-32.

Laxton D., Tatlow R. (1992), A simple multivariate filter for the measurement of potential output, Bank of Canada 59.

Marcellino M., Musso A. (2011), The reliability of real-time estimates of the euro area output gap, Economic Modelling, 28(4), 1842-1856.

Massmann M., Mitchell J., Weale M. (2003), Business cycles and turning points: a survey of statistical techniques, National Institute Economic Review, 183, 90-106.

Mitchell J. (2007), Density Estimates for Real-time Eurozone Output Gap Estimates. In Growth and Cycle in the Eurozone (eds. G.L. Mazzi and G. Savio). Palgrave, Basingstoke, pp. 310-321.

Mitchell J., Robertson D., Wright S. (2014), What univariate models tell us about multivariate macroeconomic models. University of Warwick.

Morley J.C. (2002), A state-space approach to calculating the beveridge-nelson decomposition, Economics Letters 75(1), 123-127.

Morley J.C., Nelson C.R., Zivot E. (2003), Why Are the Beveridge-Nelson and Unobserved Components Decompositions of GDP so Different? Review of Economics and Statistics, 85, 235-243.

Morley J., Piger J. (2008), Trend/cycle decomposition of regime-switching processes, Journal of Econometrics, 146, 2, 220-226.

Murray C.J. (2003), Cyclical Properties of Baxter-King Filtered Time Series, Review of Economics and Statistics, 85:2, 472-476.

Neiss K., Nelson E. (2005), Inflation dynamics, marginal cost, and the output gap: evidence from three countries, Journal of Money, Credit and Banking, 37, 1019–1045.

Nelson C.R. (2008), The Beveridge-Nelson decomposition in retrospect and prospect, Journal of Econometrics 146(2), 202 - 206.

Nelson C.R., Kang H. (1981), Spurious Periodicity in Inappropriately Detrended Time Series, Econometrica, 49(3), 741-51.

Nelson C.R., Plosser C.I. (1982), Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics 10: 139-62.

Nilsson, N., Gyomai, G. (2008), Cycle extraction: A comparison of the Phase-Average Trend method, the Hodrick-Prescott and Christiano-Fitzgerald filters, OECD.

Orphanides A., van Norden S. (2002), The Unreliability of Output-Gap Estimates in Real Time, Review of Economics and Statistics, 84, 4, 569-583.

Osborn D.R. (1995), Moving average detrending and the analysis of business cycles, Oxford Bulletin of Economics and Statistics, 57, pp. 547-58.

Pedersen T.M. (2001), The Hodrick-Prescott filter, the Slutzky effect and the distortionary effect of filters, Journal of Economic Dynamics and Control, 25, pp. 1081-101.

Perron P. (1989), The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, Econometrica, 57(6), 1361-1401.

Perron P., Wada T. (2009), Let's take a break: Trends and cycles in US real GDP, Journal of Monetary Economics, 56(6), 749-765.

Pesaran M.H., Smith R. (2006), Macroeconometric Modelling With A Global Perspective, Manchester School, 74(s1), 24-49.

Planas C., Rossi A. (2004), Can inflation data improve the real-time reliability of output gap estimates?, Journal of Applied Econometrics, 19, 1, 121-133.

Proietti T. (1997), Short-Run Dynamics in Cointegrated Systems, Oxford Bulletin of Economics and Statistics, 59, 405-22.

Proietti T. (2006), Trend-Cycle Decompositions with Correlated Components, Econometric Reviews, 25(1), 61-84.

Proietti T. (2009), On the Model-Based Interpretation of Filters and the Reliability of Trend-Cycle Estimates, Econometric Reviews, 28(1-3), 186-208.

Proietti T., Musso A. , Westermann T. (2007), Estimating potential output and the output gap for the euro area: a model-based production function approach, Empirical Economics, 33(1), 85-113.

Ravn M.O., Uhlig H. (2002), On adjusting the Hodrick-Prescott filters for the frequency of observations, Review of Economics and Statistics, 84, pp. 371-80.

Sinclair T.M. (2009), Asymmetry in the Business Cycle: Friedman's Plucking Model with Correlated Innovations, Studies in Nonlinear Dynamics & Econometrics, 14, 1.

Slutsky E. (1937), The summation of random causes as the source of cyclical processes, Econometrica, 5, pp. 105-46.

Smets F., Wouters R. (2007), Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach, American Economic Review, 97(3), 586-606.

Stock J.H., Watson M.W. (2007), Why Has U.S. Inflation Become Harder to Forecast?, Journal of Money, Credit and Banking, 39(s1), 3-33.

Stock J.H., Watson M.W. (2010), Indicators for Dating Business Cycles: Cross-History Selection and Comparisons, American Economic Review, P&P, 100, 2, pp 16-19.

Valle e Azevedo J. (2011), A multivariate bandpass filter for economic time series, Journal of the Royal Statistical Society, Series C, 60, 1, 1-30.

Valle e Azevedo J., Koopman S.J., Antonio R.(2006), Tracking the Business Cycle of the Euro Area: A Multivariate Model-Based Bandpass Filter, Journal of Business & Economic Statistics, 24, 278-290.

Vetlov I., Hlédik T., Jonsson M., Kucsera H., Pisani M. (2011), Potential output in DSGE models, Working Paper Series 1351, European Central Bank.

Wildi M. (2008), Efficient multivariate real-time filtering and cointegration. Idp-working paper, idp-wp-08sep-01, Institute of Data Analysis and Process Design, ZHAW.

Watson M.W. (1986), Univariate Detrending Methods with Stochastic Trends, Journal of Monetary Economics, 18, 49-75.

Yule G.U. (1921), On the time-correlation problem, Journal of the Royal Statistical Society, pp. 497-526.

Zarnowitz V., Ozyildirim A. (2006), Time series decomposition and measurement of business cycles, trends and growth cycles, Journal of Monetary Economics, 53(7), 1717-1739.

10

The OECD System of Composite Leading Indicators

Authors

Gyorgy Gyomai
OECD
gyorgy.gyomai@oecd.org

Roberto Astolfi
OECD
roberto.astolfi@oecd.org

Nadim Ahmad
OECD
nadim.ahmad@oecd.org

Table of Contents

10.1 Historical background

The OECD system of Composite Leading Indicators (CLIs) was first developed in the early 1970s amidst renewed interest in business-cycle research — a direct consequence of the 1969-1970 recession in developed economies. The deeper and more global recession that followed in the mid-70s reinforced the need for such a tool, leading to the creation of a dedicated OECD Working Party on Cyclical Analysis and Leading indicators in 1978. The work of this group established the broad parameters, shape, and purpose of the OECD CLIs, whose underlying structure has remained broadly unchanged since 1981.

The remainder of this chapter describes the system and the continued modifications and improvements that have been implemented since then[1]. It begins by providing some background to the purpose of the CLI system, including a discussion of its definition and scope. The CLI methodology is then presented followed by a section on the dissemination of the results along with a discussion on the CLIs revisions. The chapter ends with a set of conclusions.

10.2 Purpose

From the outset, the objective of the OECD CLIs has been to provide *"qualitative indicators of the business-cycle outlook for the short term future"*.

It is instructive to say a few additional words here regarding the qualitative nature of the CLI and what is meant by "short-term" and perhaps the most important point requiring clarification concerns the definition of the business-cycle.

10.2.1 Defining and measuring the business-cycle

From its inception, the OECD CLI system has generally taken the definition provided by Burns & Mitchell (Burns and Mitchell (1946)) as its starting point:

*"Business-cycles are a type of fluctuation found in the **aggregate economic activity** of nations that organise their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in **many economic activities** followed by similarly general recessions, contractions and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business-cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own."*

Whilst clear and precise that definition does nevertheless leave unanswered some questions, which require a little elaboration. What, for example, should be the basis for measuring or defining "economic activity"? And how should the cycle weight these *many economic activities* to arrive at an *aggregate*?

"Aggregate economic activity" clearly suggests a single measure, with real GDP, measuring all production in monetary terms in a country, being the obvious candidate. However, economic activity is not necessarily uniquely defined in this way. Alternative measures of activity in the economy could focus on many other variables, such as employment, business start-ups, wealth, international trade, etc. Like GDP these could form the singular basis for defining the business-cycle but they could also be combined in a composite weighted and normalised form to create a composite coincident indicator of the business-cycle.

In this context, it is interesting to consider how the NBER (one of the main inspirational sources of the OECD CLI system) chose to resolve the underlying issue.

---

[1] For further information see (OECD (1987)) and (Gyomai and Guidetti (2012)).

"The NBER does not define a recession in terms of two consecutive quarters of decline in real GDP. Rather, a recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales." (NBER's business-cycle Dating Committee (2010)).

In the debates leading to the creation of the OECD CLI system, both approaches were considered. Indeed, a total of 15 different ways of expressing the concept of aggregate economic activity were proposed by participating OECD member countries before finally settling on targeting real GDP "as the single most important target variable" (OECD (1987)).[2]

In practice however, at least until April 2012, the business-cycle was necessarily defined, for operational purposes, on the basis of the Index of Industrial Production (IIP). This reflected the fact that the real GDP figures needed to quantify the reference business-cycle were available on a quarterly basis for only half of the member countries (and initially none was available on a monthly basis). Instead, the IIP was available for all OECD member countries on a monthly and quarterly basis.

Of note is also the fact that the IIP represented the most "cyclical" component of GDP, accounting for approximately 35% share of gross value added in the mid 1980s. and that the industrial sector, being a significant consumer of services activities, also drove supply in a significant part of the private service sector.

Since April 2012, in response to improvements in national statistical information systems (i.e all OECD countries now produce quarterly estimates of GDP) and because of the industrial sector's diminishing share of total GDP (direct and indirect) in recent decades in most OECD economies, the CLI system switched to using GDP as the reference series (Gyomai and Fulop (2012)).

However, notwithstanding the point that the OECD CLIs now take GDP as the underlying basis for defining the business-cycle, questions remain on exactly what metric of the GDP business-cycle should be considered. Three variants are available:

- The *classical business-cycle* variant which defines business-cycles on the basis of real GDP index values. Periods of positive growth (expansion) and negative growth (classical contraction) are used to define the cycle.

- The *deviation from trend cycle* (often known as growth-cycle) which removes persistent long-term trends from the real GDP. Cyclical upturns are characterised by faster than trend growth and cyclical downturns by slower than trend growth.

- The *growth rate cycle* which is based on growth rates of the underlying reference series, with periods of accelerations and decelerations in growth.

The Methodology section below provides further details of the techniques applied in the cycle identification process.

The differences above are not mere semantics: Although the underlying business-cycles may be 'broadly' similar, the turning points are not. In particular, the growth rate cycle generally yields more turning points (and hence more cycles) than the other variants. Therefore it is an important consideration which of the three variants one should use as the target, especially because the choice of the cycle critically influences the choice of composite indicators.

The OECD CLI system focuses on the *deviation from trend* variant because this measure is generally better equipped to deal with the policy needs relating to macroeconomic management of business-cycles. Macroeconomic policies aim at reducing amplitudes in fluctuations of economic activity. With the final goal of alleviating negative welfare implications of cyclical downturns, they are markedly different from long-term structural

---

[2]This forms one of the important differences in the underlying business-cycle definition used by the OECD CLIs and NBER. However these differences seem on the surface greater than they are; the many economic variables used in the NBER system have co-incident cycles with the observed cycles in real GDP.

policies. Therefore, in order to improve the timing of policies addressed to medium-term fluctuations, a need to disentangle trends and medium-term cycles arises [3].

For completeness, and in order to fulfill specific user preferences, the OECD also calculates/derives CLIs for the "classical business-cycle" and the "growth rate cycle". It is important to note in this context that the individual indicators used in constructing the overall composite indicator for a given country are selected and optimised on the basis of the deviation from trend measure. So, although the other business-cycle variants still perform relatively well, they are not optimal.

BOX 1.: Deviation from trend cycles and output-gap.

The output-gap — a concept central to fiscal and monetary policy making — is typically defined as the difference between the actual output and potential output (or indeed, often as the difference between potential output and actual output). As the potential output is non-observable, economists use various techniques to estimate it. A simple, purely econometric technique defines potential output on the basis of trend-output. In modern macroeconomic modelling this simpler technique of estimating and projecting potential output is replaced by a *production-function approach* — an approach that requires a more detailed model of the economy, and can easily incorporate expert judgement. Note that the two approaches, although similar, can yield different output-gap estimates especially at the end of the series.

The deviation from trend business-cycles used for the CLI calculations is the sequence of output gaps obtained by the simpler, trend based approach.

10.2.2 Qualitative nature of the CLIs

The CLIs can be considered as qualitative indicators as they are optimised to anticipate the turning points in the growth-cycle and not for measuring the speed or strength of a recovery or downturn. The loss function used in the construction of CLIs gives larger weight to turning-point correspondence with the reference series, at the expense of precise tracking of the reference series outside the turning-point regions. As a result, the qualitative nature of the CLIs is a key feature to be considered when interpreting the CLIs.

Figure 10.1, representing the CLI and the estimated cycle for the OECD area, can illustrate further the qualitative nature of the CLI. The two series show strong co-movements, with the turning points of the CLI preceding those of the cycle.

The interpretation of cyclical turning points should proceed as follows. At a peak, the output gap, defined as the difference between real GDP and trend GDP, is at its maximum, which, depending on the size/ amplitude of the gap itself, may indicate for example inflationary pressures or unsustainable investment. On the other hand, at troughs the output gap is usually negative, which may indicate under-investment and most likely disinflation.

Turning points can also be interpreted in a dynamic context. A peak means that the short-term growth rate of the economy falls below the long-term trend growth (and it stays below this rate until the changes in the business-cycle values are negative), whereas a trough marks a passage from below trend growth rate to a stronger than trend growth rate.

There is a risk that one would intuitively interpret higher peaks and lower troughs as stronger/ weaker growth. As noted above, however, such conclusions may be misplaced, Whilst there is a reasonable correlation between the various CLI phases and growth rates, it does not necessarily follow that a higher peak (lower trough) necessarily means stronger (weaker) growth relative to lower peaks (higher troughs). In this sense, the appropriate interpretation of the numeric values of the CLI relates to the degree of confidence one can attach to

---

[3]The longest medium-term cycles for which cyclical macroeconomic policies can be effective are generally placed in the range of 8 to 12 years by the economic literature on business-cycles.

Figure 10.1: OECD area Composite Leading Indicator (CLI) and economic activity (long-term trend = 100)

- •Economic activity above long term trend
- •Economic activity below long term trend
- ----- •Composite Leading Indicator
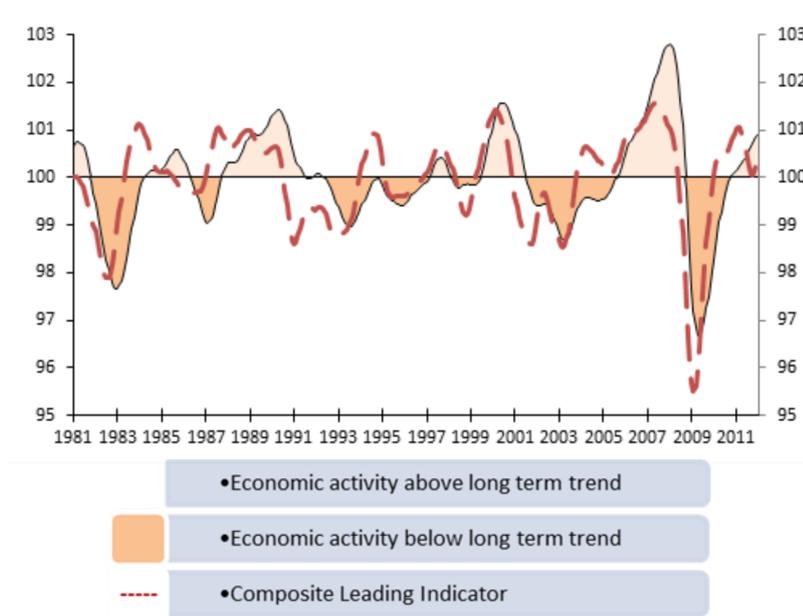
the CLI outlook - the further a peak or trough is from the long-term trend conventionally set to 100, the greater the confidence that can be attached to the CLI outlook.

10.2.3 The short-term outlook

The CLIs are designed to have a typical lead of between 6 and 9 months relative to the deviation from trend cycle turning points. In rare cases, especially for emerging economies, components yielding lead times in this range are difficult to find so the composite has slightly shorter leads.

In practice however the timeliness of the data releases affects lead times. The timeline below illustrates this more clearly. The CLI Press Release published at time T is based on the data available at that time which typically relates to $T-2$ for most CLI components. Therefore the 6 to 9 months lead of the CLI becomes a 4 to 7 months lead compared to the publication months. However, given that real GDP data and therefore direct information on the stance of the deviation from trend cycle is only available 3 to 5 months following the reference period, the CLI provides an informational lead of 7 to 12 months [4].

Figure 10.2: CLIs' Timeline

It is important also to note that, in practice, the OECD CLIs have varying lead times depending on the country and period being covered. Cycles vary, with different cycle lengths and different causes, and therefore, although there is an average tendency of the lead times, observations for a given cycle will vary from the central tendency. Nonetheless, with continuous monitoring and recalibration, the OECD attempts to maintain the lead/lag relationship in a range as narrow as possible. The OECD periodically revises the indicators (at least once every 5 years), but also strives to continue to improve the underlying econometric techniques for

---

[4]This information flow is further complicated by the fact that cycle estimates after initial publication are later revised, partly because of estimation methods incorporating new data, and partly because the data itself is revised by statistics offices. As a consequence, the "true" state of economic activity with respect to its trend can only be inferred several years after the reference period.

trend-filtering, seasonal adjustment, noise suppression methods, etc. An additional effort is made to ensure that the methodology achieves low revisions as well as stable and smooth, but still leading, indicators.

10.3 Why rely on composite indicators?

The advantage of composite leading indicators over the individual component series is that it achieves a better trade-off between responsiveness and stability. Composite indicators can be constructed to have fewer false alarms and fewer missed turning points than its individual components. Moreover, they tend to have more stable lead-times. The composites have the capacity to react to various sources of economic fluctuations, no matter whether the causes of fluctuations are endogenous or exogenous, and at the same time can be resilient to perturbations affecting only one of the components.

> BOX 2.: CLIs and conventional OECD forecastss
>
> The OECD CLI system is one of the prominent products published by the OECD to give an assessment of the short-term macroeconomic outlook. Among these products, the *Economic Outlook*, produced biannually by the Economics Directorate of the OECD, gives a comprehensive assessment of current economic developments. It also provides the economic outlook over the coming two years using quantitative forecasts for real-GDP growth and the evolution of the output-gap.
>
> Inbetween *Economic Outlook* publications, the OECD also releases an *Interim Outlook* with a more limited country scope, a stronger emphasis on the short-term (two quarters ahead), quarterly quantitative forecasts of key macroeconomic variables, and a brief analysis of recent economic developments.
>
> The OECD CLIs complement these products by providing a higher – monthly – frequency; however, as noted above, CLIs give a coarser indication of business-cycle developments. In addition, unlike the semestral and interim outlooks the CLI focuses on the assessment and does not develop policy recommendations nor does it provide elaborate commentary on the economic causes.

10.4 Methodology

10.4.1 Overview: Building blocks

Figure 10.3 presents the steps of the OECD CLIs construction and production process. The flow diagram displays the various building blocks and procedural steps taken either developing a Composite Leading Indicator for a country for the first time, or when carrying out a review of an existing indicator.

In the *Pre-selection phase* the reference indicator for the business-cycle is pinned down, and a long list of potentially leading time-series are identified.

In the following *Filtering phase* all the leading candidate series, as well as the business-cycle time-series, are manipulated to reveal their cyclical components.

In the first round of *Cyclical evaluation* the leading candidate series are tested against the deviation from trend cycle and a shorter list of promising candidates is retained. With a heuristic search several combinations of the series from the short-list are composed into competing CLIs in the *Aggregation phase* and consecutively these CLI variations are tested against the business-cycle.

A loss function combining statistical properties and practical considerations on the actual implementation is used to assign scores to each CLI variant. The loop of *Components selection, Evaluation, Aggregation* is

Figure 10.3: CLI construction workflow
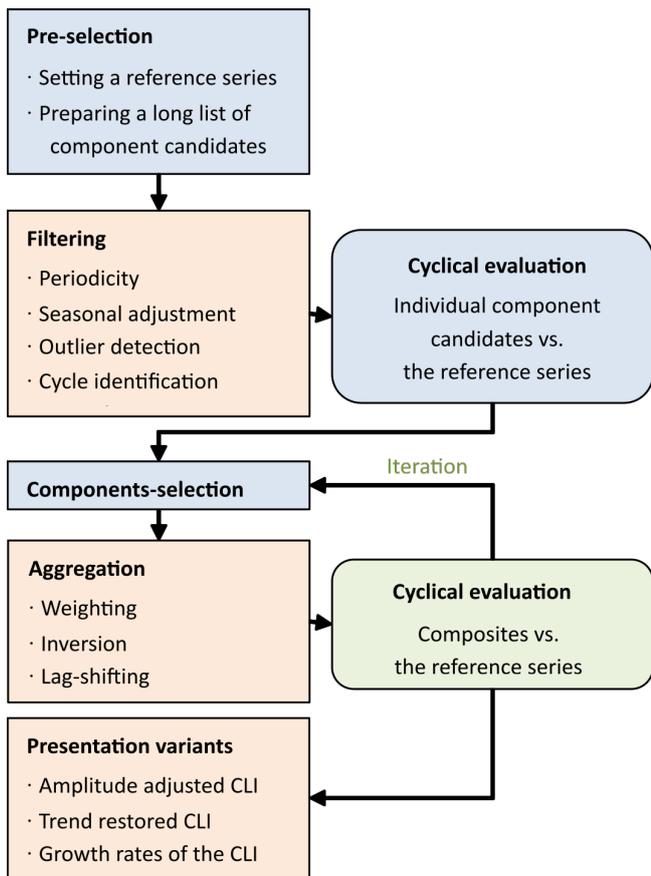
repeated until the score is minimised. Finally the best CLI composition is fixed, and *Presentation* variants are calculated.

For regular monthly production, a shorter workflow is applied, which only includes the *Filtering*, *Aggregation* and *Presentation* variants The OECD monitors the quality of the CLIs as part of the monthly production process and occasionally it intervenes by changing the seasonal adjustment and outlier detection specifications.

Reviews that include all the construction steps of a CLI occur periodically (every 2 to 5 years depending on the country) to ensure that the CLI retains its relevance. Some component series, for example, may no longer be appropriate for CLI purposes, either for economic or statistical reasons (e.g. their timeliness may worsen, and, indeed, in some cases their production may be suspended altogether). In this context it is important to note that changes to the composition of the CLIs has a retroactive effect, and historical estimates of the CLI are also replaced after a review.

The following sections provide a more detailed description of the various steps described above.

10.4.2 Pre-selection

Reference series

As noted above the OECD's preferred choice for the business-cycle reference series is the GDP. However, in the 1980s suitable GDP time-series were not readily available from many OECD economies.

Until March 2012 therefore, the OECD system of Composite Leading Indicators used the index of industrial production (IIP) as a proxy reference for the business-cycle, which was available on a monthly basis and has

also, historically at least, displayed strong co-movements with GDP.

The reliance on a proxy reference meant that the IIP replaced GDP in all operational aspects in the CLI construction work-flow. Nonetheless, when it came to business-cycle turning-point dating — which was an important element of the CLI related communication in the first two decades of the CLI system — the tentative turning-point signals in the deviation from trend cycle (derived from IIPs) were checked for robustness with the developments in quarterly GDP, whenever these latter figures were available.

In March 2012, the OECD investigated whether methods could be applied to generate monthly estimates of the deviation from trend cycle based directly on official quarterly GDP figures. This investigation has demonstrated that it was feasible to do so, whilst also continuing to provide high quality CLIs. From April 2012 therefore the OECD switched to using GDP as the reference, ceasing to rely on the IIP as an intermediate target.

Candidate component series

The process of CLI construction starts with the task of data collection and data cleaning. A long list of times-series that are likely to lead movements in aggregate economic activity are gathered for each country. The OECD primarily relies on short term economic statistics datasets collected by the OECD's Statistics Directorate covering many subject areas such as:

- GDP and its components

- industrial production and selected commodity output variables

- business and consumer tendency survey series

- selected manufacturing variables (sales, stocks, orders etc.)

- indicators from other industrial/business sectors (i.e. construction, retail/wholesale sales)

- labour market series

- consumer and producer prices

- monetary aggregates

- interest rates, exchange rates and asset prices

- international trade and balance of payments data

Beyond the regular OECD data collections the set of candidates is augmented with series collected specifically for testing in the CLI framework from national statistical offices, central banks or other data providers, often receiving help and expert advice from statisticians and economists at these institutes. The criteria for selecting candidate series are as follows:

- Economic relevance

  - *Economic significance*: the observation of a statistical leading relationship between a potential components series and the reference series is not in itself sufficient — an economic justification for the relationship is also needed.

  - *Breadth of coverage*: series with a broad coverage of economic activity are preferred to narrowly-defined series.

- Practical considerations

  - *Frequency*: monthly series are preferred to quarterly.

   – *Revision*: series that are not subject to significant revisions are preferred. For the same reason smooth series are also preferred to noisy series. Noise makes the leading signal more ambiguous and translates into revisions in the cyclical component throughthe smoothing filter which is applied to all series.

   – *Timeliness*: data should be available very soon after the end of the reference period.

   – *Length*: long time series with no breaks are preferred to short time series as the latter may hinder a robust evaluation of leading/lagging relationships due to reduced number of turning points they include

Potential leading indicators typically meet one of four types of economic rationale categories, shown below:

- *Early stage*: indicators measuring early stages of production, such as new orders, order books, construction approvals, etc.

- *Rapidly responsive*: indicators responding rapidly to changes in economic activity such as average hours worked, profits and stocks.

- *Expectation-sensitive*: indicators measuring, or sensitive to, expectations, such as stock prices, raw material prices and expectations based on business survey data concerning production or the general economic situation/climate e.g. confidence indicators.

- *Prime movers*: indicators relating to monetary policy and foreign economic developments such as money supply, terms of trade, etc.

Notwithstanding the fact that there may be a one-to-many relationship between a potential component indicator and the four rationale categories above, the CLI system attempts to balance the composition of a given CLI and the cyclical events that can impact on it, by including indicators from each of the four categories. A balanced set of indicators allows to reap the benefits of a composite indicator, as opposed to single indicators: to work for a multitude of developing cycles, and show resilience to idiosyncratic noise, i.e. uncorrelated erratic behaviour, specific to each individual series but vanishing in composite indicators.

10.4.3 Filtering

Once the candidates for a Composite Leading Indicator have been selected the second step of construction begins (corresponding to the first step in the monthly production cycle). First, quarterly time series are converted into monthly series to match the final CLI frequency. Second, using a sequence of filters, the OECD removes those factors, such as trends, seasonal patterns, outliers and noise, that may obscure the underlying cyclical patterns in the component series. Frequency conversion and approaches used to identify and remove the influence of disturbing factors are described below.

Frequency convertion

OECD CLIs are published monthly and are largely composed of monthly leading component series. However, some of the components and, starting from 2012, also the reference series, are only available quarterly. Therefore, these series need to be converted to a monthly frequency before the aggregation. Prior to the 2008 methodological review, the frequency conversion took place just before aggregation, and therefore the filters below were maintained for both quarterly and monthly time series. After the review, the frequency conversion from quarterly to monthly is the first transformation in the filtering sequence. Monthly series are achieved via linearly interpolating quarterly series and aligning them with the most appropriate month of the quarter, depending on the nature/construction of the quarterly series. For most series this is the central month of the quarter but, where quarterly data refers to the end of the quarter, the linearly interpolated series start on the last month, and for quarterly series based on surveys conducted in a given month of the quarter, the

interpolated series start on the month of the survey. The linear interpolation technique may appear simplistic, but in the context of the CLI, the methodology chosen for interpolation has little impact on the final business-cycle estimate. This is mainly because the smoothing filters in a later stage in the filter sequence dampen sub-annual variations in the time series, bringing the resolution of monthly and quarterly series closer together.

Seasonal adjustment and outlier detection

Many of the component series used in the CLI system are seasonally adjusted by the source provider, usually statistical offices. This is not the case however for all component series. In cases where only non seasonally adjusted data is provided by the source, seasonal adjustment is performed by the OECD following Eurostat's ESS guidelines on seasonal adjustment (Eurostat (2015)). However, sometimes potential residual seasonality may remain in the adjusted estimates and leak into the cyclical component identified in later stages of the filtering sequence, usually showing up as an annual cycle superimposed on the underlying business-cycle. In these cases the production process is restarted after having eliminated any residual seasonality.

Outliers are observations in component series that lie outside the normal range of expected observations. Often their cause is identifiable: for example, a strike, a change in regulation, etc. These outliers, even if identified for the sake of seasonal adjustment, often remain part of the seasonally adjusted series. However, for cyclical analysis they have to be filtered out in most cases — especially when they do not trigger a business-cycle turning-point. In cases when an extreme event triggers a whole set of changes in economic activity the rule of removing outliers is more ambiguous.

The OECD system of leading indicators utilises the TRAMO module of the TRAMO/SEATS seasonal adjust-ment program to identify outliers in each series. The OECD detects and filters out "additive" outliers (tempo-rary shocks), and analyses "level shift" outliers (permanent shocks) but does not necessarily filter these out; it is based on expert judgement whether level shifts are eliminated or not. After the location and nature of the outlier has been identified, the outliers are replaced by estimated values. In addition, TRAMO provides estimates for missing values.

Cycle identification: (de-trending and smoothing)

The OECD CLI system uses the deviation from trend approach. This means that co-movements and sim-ilarities in patterns between the reference series and the individual CLI components are evaluated via the smoothed and de-trended versions of these series. This makes the cycle extraction (the equivalent of de-trending and smoothing) a crucial step in the CLI selection and production process. The process of removing these factors can be performed in a single step (referred to as band-pass filtering), or separated into distinct steps for trend removal and smoothing.

The cycle extraction exercise can be approached slightly differently than the distinct steps of trend removal and smoothing. Instead of observing the series in the time domain, they can be regarded as a complex combination of cycles with varying cycle length, and hence the observable series is characterised by the relative phase-shifts of its constituent cycles, and even more importantly, the relative importance (amplitude) of its constituent cycles. The trend part of the series is comprised of the lowest frequency (longest cycle length) constituents, whereas the noise is formed by its high frequency (short cycle length) constituents [5].

Once the translation of the filtering problem from the time domain to the frequency domain has been made, it is easier to formulate exactly what type of de-trending and noise suppression has to be applied to the series, even if the time series manipulations are later performed in the time domain. The business-cycle of interest is usually specified in terms of the typical cycle-length of its fluctuations. The definitions are not unanimous, but vary in close ranges: the shortest cycles of interest range from 1 to 3 years, whereas the longest cycles

---

[5]See (Pollock (2008)) for a thorough introduction to the related mathematical concepts of this decomposition.

considered are between 8 and 12 years in the related macroeconomic literature [6]. In accordance with the above ranges, the OECD CLI system filters reveal/retain the cyclical constituents between 1 and 10 years of each time series.

Up until November 2008 the OECD CLI system determined the long-term trend using the Phase Average Trend method (PAT) developed by the US National Bureau of Economic Research (Boschan and Ebanks (1978)). Series smoothing was conducted using the Month for Cyclical Dominance (MCD) method. These methods were designed in the time domain, and therefore their specifications were set by rules of thumb, or reasoning different from that of a band-pass filter specification. As a result the presented cyclical components have shown slightly more noise and filtered out shorter trends than current specifications (the pass-band of PAT and the MCD smoothing was approximately between 6 months and 8 years).

The PAT method consisted of an automated or manual identification of turning points, and essentially of a stepwise linear trend fitted on three consecutive segments (the boundaries set by the turning points). The PAT results of, and revisions to, the estimated cyclical components were sensitive to the timing when new turning points were included, and turning-point inclusions significantly altered the recent history of the cyclical estimates. Beyond this turning-point definition sensitivity, the method lacked transparency — as often expert judgment was needed to identify the turning points and rules were not easy to formalise.

Following a study (Nilsson and Gyomai (2011)) undertaken in 2008 that compared the revision properties of PAT with the Hodrick-Prescott filter and the Christiano-Fidgerald filter, the OECD replaced the combined PAT/MCD approach with the Hodrick-Prescott (HP) filter. This change not only improved the stability of the cyclical estimates, but also made the CLI production process more transparent and delivered greater operational efficiency and stability.

Before the de-trending and smoothing filters are applied the OECD determines whether a series should be modelled as additive (e.g. business tendency series) or multiplicative (e.g. series expressed in nominal terms such as monetary aggregates). The multiplicative series are subject to a log-transformation, after which they can be processed in the same way as additive series.

The cycle-extraction process is sensitive towards points at the end of the time series, meaning that large revisions can occur with the arrival of new data. To lessen this flagging effect at the end of the series, the OECD CLI system uses the TRAMO module described in the above section to provide automated short horizon stabilizing forecasts (2 to 6 months ahead). These forecasts are used to extend the original data. Consecutively cycle extraction routines process the data along with the forecasted values. Finally, the HP-filter is run as a band-pass filter with parameters set: the frequency cut-off occurs at frequencies higher than 12-months and lower than 120 months.

Normalisation, re-scaling

Naturally, even after the above steps have been conducted, the different component series used in the construction of a single composite indicator will be expressed in different units or scales. As such, the various component series that are used in the construction of a CLI are first normalised. If this were not done, series with particularly marked cyclical amplitude would have undue weight in the composite indicator. The normalisation process is achieved by subtracting from filtered observations the mean of the series, and dividing this by the mean absolute deviation of the series, and, finally, by adding 100 to each observation.

---

[6]The interpretation of cycle length of interest is dependent on the business-cycle variant that serves for the identification of the cycle. A classical business-cycle, especially for high growth economies, may have fewer cyclical turning points (and hence longer cycles) than the deviation from trend cycles for the same country, and acceleration/ deceleration cycles tend to be even shorter. In the OECD CLIs context, the targeted cycle length range refers to the deviation from trend variant.

10.4.4 Evaluation

As the CLI construction work-flow shows in Figure 10.3. the evaluation stage is used repeatedly to, first assess the leading quality of the component candidates compared to the deviation from trend cycle, and secondly to assess various combinations after aggregation of the best performing series. In the evaluation stage a range of performance measures are calculated, some of them based on the raw data and some others on the identified turning points. The sub-sections below provide details of these measures.

Turning-point detection and turning-point matching

The algorithm currently used to detect turning points is a simplified version[7] of the Bry-Boschan algorithm (Bry and Boschan (1971)). It selects local minima and maxima in the cyclical part of the series, but, at the same time, enforces minimum phase length and minimum cycle length conditions, while ensuring the alternation of troughs and peaks. By default, the minimum phase length is set to 9 months, and the minimum cycle length is set 24 months.

The turning-point identification algorithm gives an indication of the location of turning-point. It has been noticed that due primarily to the fact that censor rules are enforced sequentially, the simplified Bry-Boschan algorithm does not always locate the most plausible turning points. Because of this property the review of the turning-point detection algorithm features with high priority on the list of planned methodological reviews of the cyclical indicators team in OECD. Nonetheless, the cases are rare, and do not affect strongly the calculated historical quality measures. Caution is exercised though in the presentation and interpretation of the turning points at end of the series; the formulation of the CLI outlook is not based solely on this mechanical procedure as expert judgment complements the identification of turning points.

A second algorithm compares the turning points of two series. This turning-point matching algorithm loops through the turning points of the reference series and searches in a neighbourhood of 24 months lead and 9 months lag for a corresponding turning-point (peak for a peak, trough to a trough) in the component series to be matched. If no matching turning points are found around the turning point in the reference series the algorithm marks the turning-point as missed. If only one is found it records the lead (or lag) time, whereas if multiple matches are found it selects the closest to the reference turning-point and records its lead or lag time. A further procedure checks all the remaining turning points of the component series to be matched and marks them as extra turning points.

Missing or extra cycles

Clearly, selected component indicators should not flag extra cycles or, moreover, miss any cycles compared to the reference series. Indeed, if too many extra cycles are flagged, the risk that the CLI gives false signals becomes significant. Equally, if the CLI failed to predict several cycles in the past it is unlikely to be reliable in anticipating changes in the future.

Length and consistency of the lead

Lead times are measured in months, reflecting the time that passes between matched turning points in the component and reference series. Of course lead times vary from turning-point to turning-point but the aim is to construct leading indicators whose lead times are on average between 6 to 9 months and that have relatively

---

[7]The simplified Bry-Boschan (BB) turning-point detection algorithm follows the spirit of the original BB-algorithm in maintaining minimum phase-length, cycle-length and turning-point alternation conditions in the sequence of local-extremes of a time-series. However, the simplified BB-algorithm does not employ an outlier detection, smoothing or a final turning-point re-adjustment step on the series it operates on, as these filtering steps are de-coupled from the turning-point detection algorithm and executed earlier in the filtering sequence of components of the CLIs.

small variances. To evaluate the length of leads, both mean and median leads are used, because the mean lead on its own can be strongly affected by outliers. The consistency of leads is measured by the standard deviation from the mean lead.

Cyclical conformity

If the cyclical profiles of the reference series and the component (or composite) indicator are highly correlated, the indicator will provide a signal, not only to approaching turning points, but also to developments over the whole cycle. The cross correlation function between the reference series and the candidate components (or the Composite Leading Indicator itself) provides invaluable information on cyclical conformity. The location of the peak of the cross correlation function is a good alternative indicator of average lead time. Whereas the correlation value at the peak provides a measure of how well the cyclical profiles of the indicators match, the size of correlations cannot be the only indicators used for component selection.

As a cross-check, the lead at which the highest correlation occurs should not be too different from the median lead if the composite leading indicator is to provide reliable information about approaching turning points and the evolution of the reference series.

10.4.5 Aggregation

Aggregation of component indicators is done with a view to summarise their cycle information and improve on their individual predictive quality. In the process of aggregation several questions have to be answered such as: are all component indicators equally important in predicting the cycle? what should happen when only a partial set of data is available? do all components have the same average lead? should components be synchronised by shifting them relative to each other?

The OECD CLI system has a robust way of aggregating the CLI components, practically unchanged since its conception. Nonetheless, the CLI research agenda among others, touches upon the improvement of the composition method.

As a rule of thumb, the CLI is calculated if 60% of component data are available, both for historical and current segments of the time series. A chain linking procedure ensures that changes in data availability do not introduce jumps and discontinuities in the composite indicators [8].

Weighting

The OECD CLI system has been designed to assign equal importance, and hence equal weights, to components. Although it has always been recognised that the leading properties of the components may vary over time, and therefore a more optimal weighting system is theoretically possible, the decision to rely on equal weights was motivated by the desire to keep the system transparent, easy to maintain and replicate, and to not introduce yet another source of revisions via the weighting system. Nonetheless, the OECD continues to evaluate econometric techniques that can improve the signal quality via better aggregation of the already identified leading components. Gyomai and Wildi (2012) compared the current CLI methodology with the Multivariate Direct Filter Approach (MDFA), an approach that combines in one step the cycle identification and the aggregation problem. With the multivariate direct filter, the components of the CLI receive varying weights

---

[8] The chain linking is the construction of the indicator through cumulating growth rates according to the formula below, starting at $CLI_1 = 100$

$$CLI_t = \frac{\sum_i w_i \delta_{i,t,t-1} c_{i,t}}{\sum_i w_i \delta_{i,t,t-1} c_{i,t-1}} CLI_{t-1}$$

Where w_i is the weight of indicator i (in the default setting the weights are uniform), $c_{i,t}$ is the value of the de-trended, smoothed and normalised component i at period t, and $\delta_{i,t,t-1}$ is a binary variable with value 1 if component i is available in both period t and $t-1$, and 0 otherwise. The CLI_t is only calculated for t where: $\sum_i w_i \delta_{i,t,t-1} \geq 60\%$.

for different lead times and observation periods. The study has shown that the multivariate filter considerably improves the mean square error of prediction on a 6 month horizon, partly because of the varying weights, and partly because the filter is tailormade for the composition of the CLI (unlike in the case of the univariate HP-filter).

As noted above, component indicators used in constructing country Composite Leading Indicators have equal weights. However, for zone aggregates the CLIs are weighted reflecting the size of the GDP (on purchasing power parity) of the constituent countries.

Inversion and lag-shifting

During the indicator selection, and later in the evaluation process, it may turn out that certain component candidates have a counter-cyclical, yet leading behaviour. That means that the troughs of the indicator lead the peaks of the business-cycle and its peaks anticipate the troughs of the business-cycle. This behaviour is apparent from the cross-correlogram between the component indicator and the cycle. By inverting the component indicator the turning-point matching algorithm can produce the necessary statistics for evaluation, and in this inverted form the indicator can be immediately added to the composite.

Another feature linked to the lead time of the data may be revealed during this selection phase. Some components may have large average/median lead times (beyond a year), whereas others could have a shorter, usually more stable lead times. If this is the case direct mixing of the various leading components may — under certain scenarios - weaken the signals given by both groups. One possibility in this situation is to mix the delayed values of the longer leading indicators with the shorter leading indicators, thus synchronizing their lead-times characteristics. This results in reinforced signals, but shorter lead-times, which are easier to interpret, or at least it is easier to determine the typical prediction horizon.

At the outset of the OECD CLI system the technique of delaying longer leading indicators in the composite was used in a few countries. The production system is still capable of handling time-shifted component series, but in practice all the indicators currently produced are without delayed components. The reason for this is partly practical (testing the quality of composite indicators with various delays among components multiplies the number of candidates to evaluate, when the number of possible composites is already way too large for a fully optimal selection), and partly because observation has shown that, by purely aligning median lead times, the quality of the composite does not necessarily improve.

10.5 Dissemination

10.5.1 Audience

The OECD CLIs were designed for a heterogeneous audience. In the early days the emphasis was on expert users: specialized macro-analysts, policy makers and to a lesser degree, the broad public. A secondary, but highly relevant goal was to provide a benchmark for leading indicator producers and forecasters in member state ministries, central banks and research centres, as well as for the academia.

In more recent years (in the late 2000s) the OECD targeted a wider, non-specialized audience. The monthly Press Release that forms the main dissemination tool has been significantly streamlined with the emphasis placed on a digestible and simple message. In addition, a number of explanatory notes have been produced. The attempt to convey a simpler message is also one of the reasons for the production and dissemination of variants of the business-cycle (classical, growth rate variants) along with the headline deviation from trend measure.

10.5.2 Coverage

In the late 1970s, the OECD CLIs covered 18 out of the existing 24 OECD member countries as well as a number of zone aggregates, such as the Major 7, OECD Europe, North America and OECD Total.

Since 2006, the country coverage has been extended to now include 32 OECD and 6 non-OECD countries (Brazil, China, India, Indonesia, the Russian Federation and South Africa). Further extensions are planned to also compute an aggregated CLI for G20 economies.

10.5.3 Presentation of results

The OECD CLIs are updated monthly and disseminated via various channels usually at the beginning of the second week of each month. Three variants are computed, including the amplitude adjusted, the trend-restored, and the annual growth rate of the trend restored CLIs (briefly referred to as CLI - the annual growth rate). Along with the datasets, the OECD issues a Press Release in which it gives a short commentary interpreting the latest developments in the amplitude adjusted CLIs.

The amplitude adjusted CLI vs the de-trended reference series

The amplitude adjusted is the most straightforward way to present CLIs. Obtained as rescaled average of de-trended, smoothed and normalised component series, the amplitude adjusted CLI anticipates the turning points of the de-trended and smoothed, but not normalised, reference series. This form allows for "output-gap" type interpretations for the levels of the CLI can be loosely read as the likely size of the output-gap 6 to 9 months ahead (see Box 1).

From a negative change in the CLI one can infer that real-GDP growth is likely to fall below its long term trend and, conversely, a positive change possibly anticipates real GDP growth above long term trend growth rate (with the long term trend growth rate varying largely between countries as it is higher for transition/emerging economies while it stabilises at lower level in developed economies). Two very important notes have to be made here:

First, while the location of the CLI turning points predicts fairly accurately turning points in the deviation from trend cycle (the system is tuned to maximise the accuracy of this prediction), the amplitude of the CLI matches less accurately the size of the deviations from trend throughout the business-cycle, because the emphasis in the construction process is on the former. Recall that, in the selection of components of the CLI, several measures focus on matching turning points, and it is only through the cross-correlation indicator that the overall fit throughout the business-cycle is evaluated.

Second, the "output-gap" inherent in the monthly cycle estimate does not correspond to the official OECD output-gap estimate. In the OECD CLI system, the output-gap is the difference between the smoothed real GDP and the HP-trend of the real GDP (a proxy measure for potential GDP), whereas the official OECD output-gap calculations rely on the estimation of a long term potential GDP based on a production function coupled with the expert judgment instead of the sole application of HP-trend. Recently, nonetheless, the parameter choices in the two calculations are evaluated by OECD experts to ensure that discrepancies between the two gaps are kept to a minimum.

The trend restored CLI vs the original reference series

The trend restored CLI follows from the amplitude adjusted CLI. It is the product of the trend of the reference series (trend of real GDP index) and the amplitude adjusted CLI. This transformation of the CLI facilitates analyses of trend cycles or classic business-cycles, where the business-cycle is represented by the original

real GDP figures (non de-trended, but occasionally smoothed). The trend restoration yields peaks that occur later and troughs that occur earlier in the trend restored series (at least for positive trends, which is the general situation - with a few transitory exceptions) compared to those in the amplitude adjusted series. (The ABCD turning point labelling described in Chapter [15] details phase shifts between different variants of the cycle.) As mentioned above it is important to note that the OECD turning point chronology in CLI related documents and publications will not align optimally with the turning points in the trend restored CLI, or with the turning points in the original (non-detrended) reference series.

The annual (or 12-month) growth rate of the trend-restored CLI vs similar reference series

The annual growth rates of the CLIs are calculated from the trend restored CLI. They are comparable with real-GDP growth rates, as compared to the same period in the previous year. Some analysts prefer this variant since the reference series is most often published in this format by national statistics offices. A completely coherent cyclical analysis can be built on this form of the CLI but care is needed in interpretation to avoid mixing results and messages based on different forms of the CLI. Although the cyclical patterns in the 12-month growth rate series are similar to those observed in the amplitude adjusted series (deviation from trend), peaks and troughs in the two forms have different economic meaning and therefore should be interpreted differently.

Figure 10.4: OECD area comparison between deviation from trend cycle and growth rate cycle

CLI amplitude adjusted (left axis) — CLI 12-month rate of change (right axis)
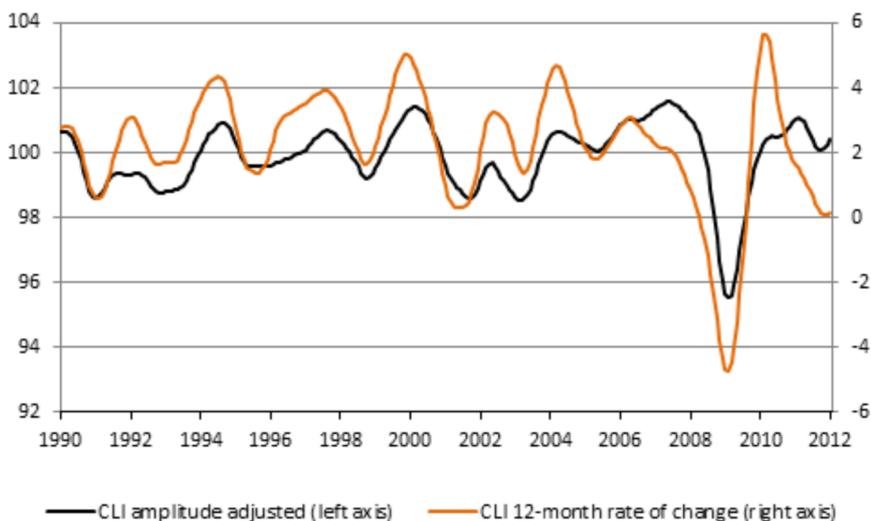
Peaks in the deviation from trend version of the reference series mean that the output-gap has started to decrease, and they also mean that the GDP growth rate falls under its long-term trend rate. Whereas peaks in the annual (or 12-month) growth rate series indicate that the growth rate starts decreasing (from its peak), the economy starts slowing/decelerating. As the peak in the growth rate cycle signals the beginning of a slowing growth, this peak is likely to happen earlier than the peak in the deviation from trend cycle which corresponds to the long-term trend growth rate. This gives the illusion of further lead-time attainable through the growth rate cycle, but it should be noted that the lead time of the growth rate series over the deviation from trend cycle is not stable, as shown in the graph above. Moreover the growth rate cycle can produce several false signals, peaks and troughs that are not followed later in the deviation from trend cycle.

10.6 Revisions

The monthly frequency imposes different requirements on the OECD CLI system than those that prevail for other, typically, bi-annual forecasting products. The message — the interpretation of the CLI based outlook — cannot change radically from one month to another, and more importantly cannot oscillate from one month to the other. Nevertheless the indicator should be able to signal radical turns when a business-cycle is approaching. The need to finally obtain a good balance between "message smoothing" and responsiveness is in the forefront of the CLI design all along the construction process (for example in the selection of candidate series), and often is decisive in selecting methodological improvement options. The revision characteristics of the CLIs are studied very closely, and this is especially the case for CLI revisions that are not due to data revision, but are rather a consequence of the inherent properties of the filters used in the production of the CLIs. In recent years a few OECD papers have analysed revisions in CLIs, and revision properties of methods applied in CLIs [(Nilsson and Guidetti (2007)), (Nilsson and Gyomai (2011)), (Gyomai and Wildi (2012))]; such analysis will be part of further attempts to improve the CLI system as a whole. To facilitate transparency and to help understanding of how turning point signals arise and evolve in real time the OECD maintains an Original Release Database which makes available consecutive monthly vintages of the CLIs since 2002 (for most countries).

10.7 Conclusion

Since its inception four decades ago, the OECD System of Composite Leading Indicators (CLIs) has established itself as one of the most followed composite indicators worldwide (Nilsson and Guidetti (2008), Astolfi et al. (2016)). Since then, the CLIs have evolved and developed continuously, partly reflecting methodological improvements and a continuous assessment of underlying leading indicators used to create the composite, but also in terms of its coverage of countries, which today includes 32 OECD Members, 6 Key Partners and several zones.

Indeed perhaps the most significant evolution over this period, at least from a dissemination perspective, has been the introduction in the late 2000s of the Press Release. Using accessible language, the Press Release aims at enlarging the CLI audience to reach the general public, alongside traditional users such as specialised macro-analysts and policy makers.

Accessibility in this respect is key. While methodological developments are crucial, as described in this chapter, interpretation of the results has always been, and remains, the most difficult challenge. Perhaps two key features are of particular importance in this context: first, the CLIs should be regarded as a qualitative/event forecast, the event being the occurrence of a turning point, and so should not be considered as a projection of the speed and strength of expansions and recessions. Second, the CLIs are optimised to minimise false signals and missed turning points, and to have stable lead times. Therefore, compared to the use of individual indicators for early detection of turning points, the CLIs present a better trade-off between responsiveness and stability.

While significant improvements have been made in recent years, the OECD cannot rest on its laurels and will therefore continue to explore refinements, tools and indeed language that improve the Composite Leading Indicators.

Bibliography

Astolfi R., Gamba M., Guidetti E., Pionnier P.A. (2016), The use of short-term indicators and survey data for predicting turning points in economic activity: A performance analysis of the OECD System of CLIs during the Great Recession. OECD Statistics Working Paper Series, 2016, OECD Publishing.

Boschan C., Ebanks W. (1978), The Phase Average Trend: A New Way of Measuring Economic Growth. Proceedings of the Business and Economic Statistics Section. Washington, D.C.: American Statistical Association.

Bry G., Boschan C. (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. Washington D.C.: NBER, Technical Paper 20.

Burns A.F., Mitchell W.C. (1946), Measuring business-cycles. New York: NBER.

Eurostat (2015), ESS Guidelines on Seasonal Adjustment. Luxembourg: Eurostat.

Gyomai G., Fulop G. (2012), Transition of the OECD CLI System to a GDP-based business-cycle Target. OECD, Statistics Directorate. Paris: OECD.

Gyomai G., Guidetti E. (2012), OECD System of Composite Leading Indicators. OECD, Statistics Directorate. Paris: OECD.

Gyomai, G., Wildi M. (2012), OECD-CLI's for G7 Countries: a Comparison of the Hodrick-Prescott Filter and the Multivariate Direct Filter Approach. Paris: OECD, Statistics Working Papers, 2012/05, http://dx.doi.org/10.1787/5k4dhnzw8xr3-en

Maravall A. (2005), An Application of the TRAMO-SEATS Automatic Procedure; Direct vs Indirect Adjustment. Madrid: Banco de Espana, Working Paper No. 0524.

NBER's business-cycle Dating Committee (2010), The National Bureau of Economic Research. Retrieved July 30, 2012, from US business-cycle Expansions and Contractions: http://www.nber.org/cycles.html

Nilsson R., Guidetti E. (2007), Current Period Performance of OECD Composite Leading Indicators (CLIs). OECD, Statistics Directorate. Paris: OECD, Statistics Working Papers 2007/1.

Nilsson R., Guidetti E. (2008), Predicting the Business Cycle. How good are early estimates of OECD Composite Leading Indicators? OECD Statistics Brief.

Nilsson R., Gyomai G. (2011), Cycle Extraction: a Comparison of the Phase-Average-Trend Method, the Hodrick-Prescott and Christiano-Fitzgerald Filters. OECD, Statistics Directorate. Paris: OECD, Statistics Working Paper No. 39.

OECD (1987), OECD Leading Indicators and business-cycles in Member Countries 1960-1985. OECD, Department of Economics and Statistics. Paris: OECD.

Pollock D.S. (2008), Statistical Fourier Analysis: Clarifications and Interpretations. Journal of Time Series Econometrics , I (1), 1-47.

U.S. Census Bureau (2011), X-12-ARIMA Reference Manual. U.S. Census Bureau, Statistical Research Division. Washington: U.S. Census Bureau.

11

The Stock and Watson Approach to Composite Indexes of CLIs

Author

Filippo Moauro
ISTAT
moauro@istat.it

Table of Contents

11.1 Introduction

The idea of summarizing the state of the general economic activity in composite indexes of coincident and leading indicators comes from the pioneering work by Mitchell and Burns (1938) and Burns and Mitchell (1946), further developed by their colleagues at the National Bureau of Economic Research (NBER). This work led to a list of coincident, leading and lagging indicators for the US whose evolution has attracted over the years economists, policy makers and practitioners interested in tracking and forecasting the state of the macroeconomic activity. Nowadays there exists a large amount of cyclical indicators also for other economies all developed in the spirit of the original contribution given by Burns and Mitchell both at institutional and at academic level.

The large volume of literature developed around these issues has generated a multitude of contributions, going from the development of new methods for selecting the indicators, the choice of the target variable and their modelling as well as a variety of examples of composite coincident and leading indexes. Marcellino (2006) provides an exhaustive review on leading indicators with a structured presentation of main features, the set of selection methods for the target and leading variables, the variety of model or non-model based methods involved in their construction and the most recent developments.

In this chapter we focus our attention on the model based approach for constructing composite cyclical indicators popularized by Stock and Watson (1989). The success of this method is that for the first time the latent concept of the "state of the economy" finds a formal definition in a probabilistic model for its effective extraction. The resulting index makes the model-based counterpart to the non-model based coincident indicator developed by the Conference Board.

This chapter aims to provide an easy-to-read introduction of the Stock and Watson methodology relative to the coincident, leading and recession composite index. In section 11.2 we provide historical elements and main motivations to the development of the work; in section 11.3 we present the coincident index, focusing on model specification, identifying assumptions and preliminary analysis, statistical treatment and main features of the empirical application; in section 11.4 we move to the leading indicator pointing more on the model derivation, applied results and variable selection; in section 11.5 we present the recession index, whereas section 11.6 is devoted to advantages and disadvantages of the Stock and Watson approach; most significant extensions, main applications outside the US and links to related literature are briefly presented in section 11.7; recent applications of interest for statistical agencies are discussed in section 11.8, also with a short presentation of the results of an application to the euro area; finally, section 11.9 presents some concluding remarks.

11.2 Main motivations and historical elements

In general, as stressed by Stock and Watson (1989) (p. 352):

"the coincident and the leading economic indexes have been widely followed in business and government for decades, yet have received surprisingly little attention from academic economists".

This scepticism of academics probably comes from the critique advanced by Koopmans (1947), according to which the original work by Burns and Mitchell is based only on empirical motivations, since all measures and computations are made with minimal assistance from the economic theory and without any study of the process behind the variables involved in the construction of the cyclical indicators; in other words it is an exercise of "measurement without theory."

Furthermore, Stock and Watson (1989) attributed this scepticism also as due to a lack of a clear explanation of what the existing coincident and leading economic indicators effectively measure. In other words, it is unclear which is the target variable for a coincident index and, consequently, which variable the coincident index leads. According to Burns and Mitchell (1946) the reference cycle is the so-called business cycle, defined as

"a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similar general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own."

From the definition above emerges how the business cycle cannot be extracted from a single series only, e.g. gross domestic product (GDP), but it requires the analysis of a range of relevant indicators of economic activity. Moreover, albeit its intuitive appeal, the definition does not reflect any quantitative content, a circumstance recognized also by Burns and Mitchell in their original work (Burns and Mitchell , 1946, p. 76).

These features motivated Stock and Watson in developing new experimental composite indexes of coincident and leading indicators. Their contribution came out for the first time in a NBER working paper dated 1988. The methodology was then refined in three other publications which appeared in 1989 with comments by (Sims , 1989), and (Zarnowitz and Braun , 1989), followed by those in the two collections which appeared in 1991 (revised version of the 1988 working paper) and 1992.

The idea behind these contributions is to present a formal probabilistic model for the construction of indexes of coincident and leading variables tracking the evolution of the business cycle.

11.3 The coincident index

11.3.1 Model formulation

The coincident index model that was proposed by Stock and Watson (1989) aims at rationalizing by a probabilistic representation the judgmental procedure that was used by the Department of Commerce (today The Conference Board) to build a coincident indicator for the US economy. The fundamental idea is to separate the dynamics which are common to a set of n coincident series X_t from the idiosyncratic component, which is specific to each series.

Then, given a vector of macroeconomic time series variables X_t observed in a sample period $t = 1, \ldots, T$ which show a contemporaneous movement with the general economy, its dynamic can be represented by a linear combination of two basic stochastic elements: the former a common latent scalar variable C_t denoted composite coincident indicator CCI or more simply "index" with a linear structure modelled contemporaneously to all the variables; the latter a n-vector component of idiosyncratic movements to each elementary time series of X_t denoted as v_t and also characterized by a linear form. The model is the following:

$$X_t = \tilde{\beta} + \gamma(L)C_t + v_t, \tag{11.1}$$

$$\tilde{\varphi}(L)C_t = \delta + \eta_t, \tag{11.2}$$

$$\tilde{D}(L)v_t = \varepsilon_t, \tag{11.3}$$

where L denotes the lag operator such that $L(X_t) = X_{t-1}$, $\tilde{\varphi}(L)$ a scalar autoregressive operator such that

$$\tilde{\varphi}(L) = 1 - \tilde{\varphi}_1 L - \tilde{\varphi}_2 L^2 - \ldots,$$

$\tilde{D}(L)$ a matrix lag polynomial such that

$$\tilde{D}(L) = 1 - \tilde{D}_1 L - \tilde{D}_2 L^2 - \ldots,$$

with $\tilde{D}_1, \tilde{D}_2, \ldots$ square matrices of autoregressive coefficients, $\varphi(L)$ an n-vector of lag polynomial weights or factor loadings, $\tilde{\beta}$ a vector of constant elements, ε_t a vector of disturbances, δ a constant scalar and η_t a scalar disturbance.

The model 11.1—11.3 corresponds to the level specification of the single-index model presented in Stock and Watson (1991) (p. 66), reflecting the empirical consideration that most of time series taken in their original values or after logarithmic transformation show a stochastic trend. From a theoretical standpoint when each component series of X_t show a trend, the model above lets open the possibility that the index C_t contain itself a trend. This implies a common trend in the set of time series X_t or, according to the definition by Engle and Granger (1987), that X_t is co-integrated of order $(1, 1)$.

When by contrast data are supposed to be not co-integrated or after a preliminary test analysis the hypothesis on the presence of common trends is rejected, the model (1)-(3) can be modified to cast data transformed in simple or logarithmic differences. Then, the model is built on the transformed set $Y_t = \Delta X_t = X_t - X_{t-1}$ taking the form

$$Y_t = \beta + \gamma(L)\Delta C_t + u_t, \tag{11.4}$$

$$\varphi(L) = \Delta C_t = \delta + \eta_t, \tag{11.5}$$

$$D(L)u_t = \varepsilon_t, \tag{11.6}$$

where $\varphi(L)$ and $D(L)$ are stationary autoregressive polynomials of finite orders resulting, respectively, from the factorization of $\tilde{\varphi}(L)$ and $\tilde{D}(L)$ into $\tilde{\varphi}(L) = \varphi(L)\Delta$ and $\tilde{D}(L) = D(L)\Delta$, with Δ the difference operator such that $\Delta = 1 - L$; moreover β takes the meaning of drifts in Y_t when different from zero.

Relevant interpretation issues around the single index model 11.4—11.6 are the following:

1. It exists a single causal source of common variation among the variables Y_t;

2. The impulse response from η_t to Y_t is proportional across observable series;

3. There are multiple source of economic fluctuations, but they have proportional dynamic effects on the real variables Y_t;

4. The model provides a direct measure of the contribution of each coincident indicator in Y_t to the single index ΔC_t, simply evaluating the response of ΔC_t to a unit impulse in Y_t;

5. Standard errors around ΔC_t can be computed through the Kalman filter and smoothing technology.

11.3.2 Identifying assumptions

The single-index model 11.4—11.6 requires a set of identifying assumptions for allowing estimation of un-known model parameters and unobserved values of ΔC_t. Those presented in the original Stock and Watson (1989), Stock and Watson (1991) contributions are the following: the first concerns the property that the co-movements of the multiple set of time series Y_t arise from the single latent variable ΔC_t, implying that $D(L)$ is diagonal and the disturbances are mutually uncorrelated. More formally that,

$$D(L) = diag(d_1(L), \ldots, d_n(L))$$

$$Cov(\eta_t, \varepsilon_t) = \Sigma = diag(\sigma_\eta^2, \sigma_{\varepsilon_1}^2, \ldots, \sigma_{\varepsilon_n}^2).$$

The second assumption concerns ΔC_t, whose scale is identified restricting the variance of its disturbance term η_t to unity notably $\sigma_\eta^2 = 1$. Finally it is required that the mean of ΔC_t is a weighted average of the growth rates of the set of data Y_t, which means restricting the drift δ of equation 11.5 on the base of observed data. Several variants of the identifying assumptions proposed in origin by Stock and Watson appeared in the advancements of the literature on this topic. Very often it is imposed a zero drift in equation 11.5 like in Proietti and Moauro (2006) and Frale et al. (2011), or data are preliminary demeaned as in Mariano and Murusawa (2003).

11.3.3 Preliminary analysis

In the original application by Stock and Watson (1989), data entering in the model were subject to a preliminary analysis to test the presence of stochastic trends in any of the single series of the model and, if this is the case, whether these trends are in common within each other's.

The presence of stochastic trends in a univariate time series has a direct implication in the type of transformation to apply to gain stationarity in a series. When an autoregressive integrated moving average (ARIMA) models is adopted it traduces either in the number of subsequent differences to apply to the series to render it stationary, also referred as "order of integration" $I(d)$, or the number or unit roots in the autoregressive polynomial of the ARIMA representation.

The unit root test used originally in Stock and Watson (1989) is that by Dickey and Fuller (1979). Here it is tested for 1 versus 0 unit roots in a univariate autoregressive framework or, in other words, that the series is non-stationary against the alternative that it is stationary. Then, these tests are designed for rather different situations and become, since the beginning, very popular among practitioners, analysts and econometricians. A further advantage is that these tests are easy to implement and quite convenient from a computational standpoint.

Concerning the presence of common trends the testing procedure was originally that suggested by Stock and Watson (1988a). Here it is tested the null hypothesis that the set of n time series X_t assumed to be I(1) has $k \leq n$ common stochastic trends, against the alternative that it has $m < n$ common trends. The test is based on the roots of the estimated autoregressive representation of the time series.

11.3.4 Statistical treatment

The state space methodology is required to produce the estimates of unknown model parameters and the single-index. See Harvey (1989) and Durbin and Koopman (2013) for full details. First the model 11.4—11.6 is cast in an appropriate state space form consisting of measurement and a state equation; then the Kalman filter is run for computing the log-likelihood function through prediction error decomposition; thirdly, unknown parameters are estimated maximizing the log-likelihood function via a quasi-Newton method; finally the single-index is computed conditional to log-likelihood estimates of model parameters and the data set. The state space methodology allows from one side the estimation of ΔC_t in real time (filtered estimates) denoted here as $\Delta \hat{C}_{t/t}$, meaning that these estimates are conditional to the information set until time t; from the other the so called smoothed estimates, denoted as $\Delta \hat{C}_{t/T}$ and differing from the estimate $\Delta \hat{C}_{t/t}$ since conditional to the entire information set. Furthermore associated with this couple of estimates the Kalman

filter also computes the relative variance estimates which allows to have a measure of the uncertainty around $\Delta \hat{C}_{t/t}$ and $\Delta \hat{C}_{t/T}$.

11.3.5 Empirical application

The first application of the single index model was referred to the US economy adopting the same four-time series used by the U.S. Department of Commerce (DOC) to build its composite coincident index (nowadays the same index is published by The Conference Board). In detail data span the period from February 1959 to December 1987 and concern Δ-log of the seasonally adjusted version of the following coincident indicators:

- Index of industrial production

- Employees on non-agricultural payrolls

- Total manufacturing and trade sales in 1982 dollars

- Total personal income less transfer payments in 1982 dollars

DF and AFD-tests for the presence of a unit root did not reject the hypothesis that each of the coincident indicators were non-stationary. Further, the Stock and Watson (1988b) test for the presence of co-integration provided evidence that the series were not co-integrated.

In a subsequent application the series of employees were substituted by employee-hours in non-agricultural establishment. Results and comparison with the former application were presented in Stock and Watson (1989).

Model specification led to a AR(2) specification for the coincident index equation 11.5, as well as all the four idiosyncratic components of equation 11.6. The difference of the first model including the number of employees instead of employee-hours resulted only in one occurrence; in the former the coincident index ΔC_t enters contemporaneously to all the four time series, meaning that the vector lag polynomial $\gamma(L) = \gamma_0$ in equation 11.4; whereas in the latter the form for ΔC_t is AR(3) for employee-hours, remaining AR(2) for all the other 3 component series.

11.4 The leading index

11.4.1 Deriving the model

The construction of an experimental leading economic index was approached by Stock and Watson (1989) as an estimate of the growth rate of the coincident index $C_{t/t}$ over the next 6 months and denoted as LEI. Then, in terms of the notation presented above, the LEI is defined as $C_{t+6/t} - C_{t/t}$.

The aim was also in this case to rationalize the original NBER view according to which an economic leading indicator should provide a measure with the feature to anticipate the reference cycle by several months. Thus if the underlying idea is that it exists a common force which drives the evolution of several economic phenomena, then an autoregressive form built on the log-differences of the CCI can be exploited for obtaining predictions of the same index some steps ahead.

Notice, however, a difference among the Stock and Watson' view and the classical NBER approach: in the former a leading indicator is a measure in growth rates whereas in the latter it is expression of levels. This depends only by the empirical evidence that in the Stock and Watson' view one is more interested in the relative growth than in the absolute levels of the economic activity. Moreover the link between levels and growth rates is immediate by considering that the sum of LEI and the CCI is in fact $C_{t+6/t}$, i.e. the forecast six months ahead of the log-level of the CEI.

Similarly to the CEI, the Stock and Watson proposal for the LEI is a model based indicator. In particular a set of leading variables grouped into a vector Z_t and the index C_t are modelled through the following vector autoregressive VAR system.

$$\Delta C_t = \mu_C + \lambda_{CC}(L)\Delta C_{t-1} + \lambda_{CZ}(L)\Delta Z_{t-1} + v_{Ct}, \tag{11.1}$$

$$Z_t = \mu_Z + \lambda_{ZC}(L)\Delta C_{t-1} + \lambda_{ZZ}(L)\Delta Z_{t-1} + v_{Zt}, \tag{11.2}$$

where (v_{Ct}, v_{Zt}) are serially uncorrelated white noises, λ_{CC}, $\lambda_{CZ}(L)$, $\lambda_{ZC}(L)$, $\lambda_{ZZ}(L)$ are autoregressive lag polynomials whose order are determined empirically using a quite complicated statistical procedure. The leading variable in Z_t are first subject to transformation to render them stationary if necessary.

The VAR 11.1—11.2 is nonstandard in the sense that from one side it features C_t as unobserved and from the other it requires to be restricted in order to control the high number of parameters to be estimated. Stock and Watson (1989) deleted the higher lags of the variables in all equations of the system except the equation for the coincident variable C_t.

The joint statistical treatment of the coincident and the leading models can be carried out casting together equations 11.4, 11.5, 11.1 and 11.2 into a state space form and running the usual Kalman filter and smoothing routines for maximum likelihood estimation of unknown parameters and the involved latent variables.

In the original formulation Stock and Watson (1989) proposed a simplified procedure in two steps where first unknown parameters of model 11.4—11.6 were obtained and then the leading model 11.1—11.2 estimated conditional on estimates of the former. This procedure has the advantage that the coincident index $C_{t/t}$ is consistent with the complete system 11.4, 11.5, 11.1 and 11.2 and prevents from potential mis-specification in 11.1 and 11.2. From the other side it has the risk to produce consistent but inefficient estimates with respect to the full model when this latter is correctly specified.

11.4.2 The leading index for the US economy

The experimental LEI relative to the US economy was presented in Stock and Watson (1989) (pp. 363-376).

Variable selection and model specification of the model 11.1—11.2 resulted from an accurate multi-step statistical procedure of leading variables. A summary description is presented in this section, leaving the reader interested in all the details to Stock and Watson (1989). The empirical application needed the choice of variables entering the model, eventual transformation and smoothing of selected variables and determination of number of lags to use in the ΔC_t equation.

Starting point was a list of around 280 series split in 10 groups:

1. measures of output,

2. consumption and sales,

3. inventories and orders,

4. money and credit,

5. interest rates and prices,

6. exchange rates and foreign trade,

7. employment, earnings and labour force data,

8. wages and process,

9. government fiscal data,

10. other leading indicators from Business Condition Digest.

The list took into account series with expectational components in order to contain data able to rapidly react to eventual shocks to the economy. The screening was carried out looking at the bivariate relation to the DOC coincident indicators (coherence and phase lead between each series and the growth of DOC components), as well as the ability of any single series to Granger-cause DOC data.

This screening led to 55 series from which the final leading variables were short-listed through a stepwise regression procedure. At this stage and differently from the traditional NBER approach, selection was carried out pointing on the multivariate rather the bivariate predictive content of each of the series. These regressions involved the six-month growth $C_{t+6/t} - C_{t/t}$ as dependent variable on current and past values of candidate leading series. Several leading indexes were constructed changing each time the base set of regressors by including twelve lags of each of the candidate trial variables in the six-step ahead regression; these were ranked according to a criterion based on the R^2 statistic. The series with the greatest value of the criterion function was added to the index and the procedure was repeated until the desired number of variables was added. The final set was chosen by considering those series that most often were found in the final index, starting from different sets of base variables.

Controlling high frequency noise of growth rates of some leading variables was the second order of problems in the determination of the final form of the leading index model. The solution was found by applying the smoothing filter $S(L) = 1 + 2L + 2L^2 + L^3$, the same as that used for noisy series by the DOC until a benchmark revision undertaken in 1989. The application of a smoothing filter was also a way to reduce the number of estimated model coefficients since one can let the leading indicators enter in the regressions with a lower number of lags.

The last problem in computing the LEI index was finding the number of lags of each series in the ΔC_t equation of the LEI model or, in other words, the order of $\lambda_{CZ}(L)$ in equation 11.1. For this purpose the Akaike information criterion (AIC) was used in a regression of $\Delta C_{t/t}$ on four lags of $\Delta C_{t/t}$ and the selected leading variables taken at only 1, 3, 6, or 9 lags for restricting the number of specifications.

The resulting LEI model specification of equations 11.1—11.2 included as leading variables the following 7 series:

1. Building permits in levels taken at 6 lags;

2. Manufacturers' unfilled orders in smoothed growth rates at 3 lags;

3. Trade-weighted nominal exchange rates between US, and the UK, West Germany, France, Italy and Japan in smoothed growth rates at 3 lags;

4. Part-time work in non-agricultural industries because of slack work in smoothed growth rates at 3 lags;

5. Yield on constant-maturity portfolio of 10 years US Treasury bonds in smoothed differences at 3 lags;

6. Spread between interest rate on 6 month corporate paper and the interest rate on 6 month US Treasury bills in levels at 6 lags;

7. Spread between the yield on constant-maturity portfolio of 10 year US T-bonds and the yield on 1 year US T-bonds in levels at 3 lags.

11.5 The recession index

The presentation of the recession index by Stock and Watson (1989) (p. 356) is interestingly preceded by a discussion on the role of identifying and forecasting cyclical turning points and the debate on this issue among

academic economists.

The debate concerned from one side the traditional view of business cycle analysts according to which contractions and recoveries are part of the macroeconomic process observed through the data, even if the behaviour of the agents is represented by distinct functions reflecting the asymmetry in their behaviour through recessionary and expansionary periods. Then, according to this standpoint the cyclical behaviour of the economy has an "intrinsic" interpretation.

From the other side, a group of economists supported the idea that there was neither nothing of intrinsic in the macroeconomic cyclical evolution, nor any shifts in the economic behaviour of the agents; rather expansions and contractions are "the results of a stable structure adapting to random shocks" (Stock and Watson (1989), p. 357). Therefore, in this case it prevails an "extrinsic" interpretation for the recession and the expansions governing the economy. This view is contained in both the Stock and Watson' coincident and leading indexes of previous sections and it is also consistent with their proposal for the recession index. This latter index is defined as "an estimate of the probability that the economy will be in a recession six months hence".

The recession index has the same informational content of the leading index and it is based on the classical Burns and Mitchell's definition of shortest period for a recession or an expansion identified in 6 months. Therefore, the choice to compute the recession index as the contraction of the coincident index for a period at least of six months.

The measure of the recession probability is approached by estimating several binary logit models and selecting from this set the most appealing in a statistical sense. The dependent variable of this model is the 0-1 monthly variable R_{t+6} where R_t is the monthly variable assuming the value of 1 if the economy is in recession according to the NBER official dating and 0 elsewhere. The regressors are given by the forecasts $\hat{f}_t(1), \ldots, \hat{f}_t(6)$ computed from a mixed autoregressive system with $\hat{C}_{t/t}$ and the following 7 leading variables:

1. manufacturing and trade inventories,

2. manufacturers' unfilled orders,

3. building permits,

4. yield on constant-maturity portfolio of 10 years US Treasury bonds,

5. spread between the ten-year bond yield and the interest rate on ninety-day T-bills,

6. 6-month interest rate of commercial paper,

7. industrial production of durable consumer goods.

Further analysis to this base model was also carried out including additional leading variables as regressors both one by one or in groups. The system was estimated by maximum likelihood for convenience, even if recognized that the likelihood was mis-specified since the error terms should include a moving average term. The sample period was January 1961- December 1987.

The results showed that the logit model considering the set of 6 forecasts as regressors was able to substantially reduce the unexplained variance of the binary recession variable R_{t+6}. Further, that despite expansionary phases can be predicted accurately, probabilities to forecast recessions 6 months ahead are around two third. Finally that the addition of further leading variables, in particular financial variables, was found of utility in increasing the forecast accuracy.

11.6 Advantages and disadvantages of the Stock and Watson approach

Main advantage of the Stock and Watson methodology is that it formalizes in a statistical sense the Burns and Mitchell ideas for the construction of coincident and leading indicators. It overcomes the empirical and classical procedure of constructing such composite indicators step by step, starting from the selection of the component series, their transformation to get data adherent to the definition of cycle we want to represent (typically classical cycle or deviation cycle), removal of undesired features from original data like seasonality and outliers, standardization to render the amplitude of cyclical swings present in the component series similar to each other's and, finally, their aggregation according to a simple weighting scheme. We make specific reference to the method adopted by The Conference Board (previously by the Department of Commerce).

The step-by-step approach by The Conference Board differs from that by Stock and Watson for the following elements:

1. in the former any explicit reference to a target variable (like GDP, industrial production, employment) is taken into account, whereas in the latter a probabilistic model is formalized around the common cyclical behaviour of coincident indicators for extracting the composite index;

2. in The Conference Board approach the weighting scheme is arbitrary chosen to be equal for all the components and taken invariant over time, whereas the model-based approach by Stock and Watson allows that these weights are endogenous estimated and they could implicitly vary over time;

3. lagged values of coincident indicators are not taken into account in The Conference Board approach, whereas in the latter autoregressive forms are assumed for each coincident indicator in the multivariate setup;

4. standard errors around the estimated composite index can be computed in the Stock and Watson framework, whereas it is not possible in that of The Conference Board;

5. the single index model is the result of complicated statistical methodology, whereas the procedure of The Conference Board is easier to apply with advantages also in the interpretability of the derived composite indicator;

6. in the empirical application to the US economy over the period starting from January 1959, the two indexes show a very similar behaviour in terms of dynamics, amplitude of the cyclical swings and pattern of peaks and troughs.

The Stock and Watson method is a model based procedure which recognizes the idea that the forces responsible for the cyclical dynamic of the economy are a few and that the idiosyncratic components are uncorrelated across the variables under analysis. Other advantages are more technical and derive from the use of the state space technology:

- first it is possible to compute estimates of the coincident and leading indexes both in real time and conditional to the full information set;

- second, it is possible to have an endogenous measure of the contribution of each component series;

- third, model estimate and latent composite indexes can be efficiently computed also in presence of missing values at the end of the sample for some component series. This occurrence happens when some indicators are released with delay;

- fourth, the model can be adapted to allow for a unified treatment of data revisions;

- finally, the method allows evaluating standard errors around both the coincident and the leading indicators.

Main disadvantage is its complexity with respect to non-model based procedures which, by contrast, are both easy to build and to explain, also offering an immediate interpretation.

Finally, we provide a synthesis of the comments to the 1989 Stock and Watson paper by Sims and Zarnowitz and Braun. According to Sims, the Stock and Watson method does not cope with a model with time-varying coefficients and estimated by Bayesian methods. In this respect, samples of observed time series are too short and the statistical theory is built under the assumption that samples tend to infinity. Furthermore, we do not have neither a precise knowledge of the more suitable model specification for the data at hand, nor an idea for the values of model parameters. Then, it could be more convenient the adoption of an explicit Bayesian approach which could address model parameter estimates towards sensible a-priori guesses. A time varying context could overcome the limit of assuming a fixed stochastic structure for modelling a reality which changes over time both for changes in economic policy and "because of shifts in populations, technology, tastes and resource availability" (Sims (1989), p. 394).

According to Zarnowitz and Braun (1989), they advanced the doubt that the cause of business cycle fluctuations could be more than one. In other words, modelling the cycle under the assumption of a single force governing the cycle is convenient but it is not proven and, by contrast, there is evidence that the sources of fluctuations are more than one and subject to changes over time. Concerning the leading index the more substantive critique concerned the low number of leading indicators included in the final unobserved VAR model. The selection procedure is not clear enough and conditioned by the modelling approach for which controlling the dimensional aspects is a real need.

11.7 Extensions, application and link to related literature

The work by Stock and Watson for the construction of coincident and leading composite indicators has produced a significant number of subsequent research. Amongst others, most significant contributions have been:

- that developed by Crone and Clayton-Matthews (2004) for the production of coincident indexes for the 50 US states using the same original Stock and Watson methodology ;

- the contribution by Hamilton (1989) developing a non-linear probabilistic approach able to handle to model parameters to vary over the time. In particular it applied a Markov switching modelling approach to business cycle analysis according to which the growth rates of the variables might depend on the occurrence that the economy is in an expansionary or in a recessionary phase. This approach fully addresses the critique by Sims (1989) quoted in previous Section and also adapts to the case the status of the business cycle are more than one;

- the contribution by Diebold and Rudebusch (1996) synthesizing in one model both the dynamic factor approach by Stock and Watson and the regime-switching based one proposed by Hamilton. In this way the model addresses both co-movement and non-linearity of the business cycle as advocated by the pioneering work of Burns and Mitchell;

- the contribution by Kim and Nelson (1998) who adopted the model by Diebold and Rudebusch, estimating model parameters and latent variables by means of the Gibbs sampling methodology. This latter method overcomes the approximation required using the maximum-likelihood estimation required under the proposal by Diebold and Rudebusch.

- the alternative factor based methodology proposed by Forni et al. (2000), Forni et al. (2001) for the derivation of a composite coincident indicator of the Euro area further developed by Altissimo et al. (2010) and Forni et al. (2005). The main result of this literature is the current production of the monthly

indicator Eurocoin (see www.cepr.org) able to depict the euro area business cycle. The target here is given by the underlying quarter on quarter growth of euro area GDP.

For a more extensive survey on main developments of dynamic factor models we refer to Marcellino (2006) and Stock and Watson (2006).

11.7.1 Empirical applications outside US

The Stock and Watson methodology has also seen applications outside the US. A not exhaustive list of these applications with short distinctive elements is the following:

- the contribution for Canada by Gaudreault et al. (2003) for developing a new system of composite indexes of coincident and leading economic indicators. The authors also introduced a modification to the original Stock and Watson recession index for the US economy based on logit models;

- for Lithuania by Reklaite (2011) where a complete system of coincident, leading and recession indexes is derived;

- for Colombia by Nieto (2003): In this work the author summarizes the contributions of a wider project for deriving coincident and leading indexes also with methodological modifications to the original Stock and Watson approach;

- for Greece by Hall and Zonzilos (2003) where the coincident indicator is obtained by a model which is calibrated rather than estimated. Here it is stressed the utility of the indicator for short run policy analysis;

- for Hong Kong by Gerlach and Yiu (2004) with the objective of producing current-quarter estimates of economic activity. Here The Hang Seng index, a residential property price index, retail sales and total exports are used as coincident indicators in the model;

- for Iceland by Eklund (2007), with the aim of forecasting the future state of the business cycle. Interestingly, it is developed by the author a simple approach to estimate recession and expansion probabilities based on bootstrap resampling techniques;

- for Ireland by Murphy (2005) in which the aim is to provide an economic indicator for Ireland capable of being updated on a timelier basis than estimates of GDP;

- for Japan by Fukuda and Onodera (2001) where the principal goal is to derive a set of indicators for forecasting short-run economic fluctuations;

- for New Zealand by Claus and Claus (2007) for constructing seven leading indexes of employment and comparing their forecasting performances;

- for Turkey by Akkoyun and Gunay (2012) who used an extended version of the Stock and Watson coincident indicator model to deal mixed frequency time series, ragged ends and missing data. The aim is to provide early estimates of the Turkish GDP growth.

11.8 Main applications by statistical agencies

In this section we focus on the main developments of the Stock and Watson approach of interest for statistical agencies. We refer in particular to the contributions by Mariano and Murusawa (2003), Proietti and Moauro (2006) and Frale et al. (2010, 2011) and the most recent application towards multi-factor models developed by Grassi et al. (2013). Finally the results of a specific application to the euro area are presented. The

application is focused on a graphical comparative analysis of the Stock and Watson coincident index resulting from the application with that of the Conference Board.

The interest by statistical agencies in the Stock and Watson approach derives from the need to answer by requests coming from economists, practitioners and the general public of reliable and timely short term macroeconomic indicators. In particular this method, as well as other recent methodological advances in econometrics, have been applied to increase length, coverage and timeliness of short term statistics in situations like with the production of monthly data where only quarterly indicators are available.

In a context of data production, most of the problems specific to the construction of composite cyclical indicators are simplified under several respects: first the problem of choosing the target variable, most of the time adherent to a concept of measuring something really observable (level of activity, number of persons employed, consumption); second, by the need to follow codified rules in the compilation of the variable under interest and the choice of the weighting scheme. GDP is the typical example being a top variable compiled following a bottom-up strategy according to which it results either by the sum of value added components from the supply side, or by the sum of demand components given by household and government consumption expenditure, gross fixed capital formation, net exports and change of inventories.

11.8.1 Extensions towards mixed frequency models

The extension of the Stock and Watson approach towards mixed frequency methods has been motivated by the possibility to open the range of potential coincident indicators to relevant quarterly series. The primary candidate is real GDP which represents the most important quarterly coincident indicator even if subject to greater revisions than the original four monthly coincident series.

This consideration motivated Mariano and Murusawa (2003) to extend the single-index model with the inclusion of quarterly real GDP growth; they proposed a linear state space model at the monthly observation frequency that entertains the presence of an aggregated flow. They explicitly formulated their model in terms of the logarithmic changes in the variables, however, without taking into account the non-linear nature of the temporal aggregation constraint.

The extended Mariano and Murusawa (2003) version produced very similar results to the original formulation of the coincident index in terms of position in time of turning points over the estimated sample period (January 1959 December 2000). By contrast the two versions diverged when looking at amplitude of the cyclical swings: concerning the peaks that of July 1981 is higher than that in January 1980, whereas it is the contrary in the original formulation; concerning the troughs July 1980 is deeper than November 1982 according to the new approach whereas it is again the opposite for the original formulation. The conclusion is that modelling together GDP to the original set of four indicators produces a more appealing coincident index with an economic interpretation and whose levels depict an evolution closer to the real economic dynamics.

The proposal by Proietti and Moauro (2006) moves along the same motivations than those by Mariano and Murusawa (2003) but with some relevant differences: first and foremost, the problem of modelling time series with different frequencies of observations and subject to a non-linear temporal aggregation constraint, induced by the logarithmic transformation, was explicitly afforded in the Stock and Watson setup. The proposed solution was grounded in the theory that was developed by Fahrmeir (1992) and Durbin and Koopman (2013) and requires matching the conditional mode of the states of the non-linear and the linear approximation; this operation is performed by iterating on the Kalman filter and smoother estimating equations; secondly, the model was set up in the log-levels of the variables of equations (1)-(3) rather than in the changes in their logarithms. The advantages of this formulation are twofold: firstly the mean-square errors of the estimated coincident index are immediately available both in real-time (filtering) and after processing the full available sample (smoothing). Secondly, the treatment of the aggregation constraint in the log-levels is more transparent and efficient from the computational standpoint, in that it leads to a reduced state vector dimension. Finally, the method allows to compute monthly GDP estimates that are consistent with the quarterly totals. In

the US application the sample was longer, covering the period January 1959-February 2003. Interestingly, for economic interpretation, the index of coincident indicators was computed rescaling the common factor with the factor loading of GDP and setting the drift to that of "latent" monthly GDP.

The relevance of GDP and the need to make it available at monthly frequency was at the base of the construction of EUROMIND, i.e. the euro area monthly indicator of economic conditions by Frale et al. (2011). According to this contribution a small scale dynamic factor model of the original Stock and Watson type was adopted to temporal disaggregate the main quarterly components of GDP from the supply and the demand side. The monthly indicators used in the approach were industrial production, retail turnover, number of passengers, employment, hours worked and others used to monthly estimate six value added components of GDP. The observed quarterly value-added series were distributed over the 3 months ensuring that the sum of the three distributed values is consistent with the quarterly figure. The same approach was followed to estimate GDP components from the expenditure side, by using related monthly indicators of the final demand. Finally, the estimates of total GDP were reconciled by combining the supply and expenditure side's estimates using optimal weights, which reflect their relative precision.

Most of factor model based methods for the construction of coincident indicators employ a large information set. It is also the case of EUROMIND which, however, splits it into smaller data sets on the base of the available decomposition of GDP. Then rather than using a single model for the entire data set, EUROMIND is built applying factor models of small size to each sub-data-set. Such a strategy allows a more accurate selection of the relevant monthly indicators and permits more easily understanding of the role of each demand or supply components for the evolution of overall GDP.

11.8.2 Extensions towards multi-factor models

A first extension towards multi-factor models is that proposed by Frale et al. (2010). This contribution is a variant of EUROMIND, denoted later as EUROMIND-S, with the aim of exploiting more carefully the contribution of business and consumer surveys in a two-factor model. The introduction of a second common factor in the Stock and Watson model captures the contribution of survey variables as coincident indicators, with the first factor which allows for low frequency cycles. The evaluation of the in-sample contribution of the second survey-based factor and the short term forecasting performance of the model was addressed in a pseudo-real time context. Main findings were that for the monthly estimation of quarterly euro area GDP, the survey-based factor turned out to be statistically significant for Industry and Exports. Moreover, the forecasting performance of the survey-based model was in general substantially better than that of a single factor model and more traditional autoregressive distributed lag (ADL) specifications. By contrast, the historical evolution of this variant was extremely close to that of the basic EUROMIND.

A last contribution of this review is the most recent variant of EUROMIND by Grassi et al. (2013) which allows the simultaneous calculation of monthly indicators of the economic activity for the Euro Area and its largest member states named EUROMIND-C. It is based on a parametric large scale factor model handling a very large set of time series, with mixed frequency, and subject to missing values. The construction of EuroMInd-C is based on more than 100 monthly time series and 55 quarterly national accounts series. The latter concerns the decomposition of gross domestic product according to the output and expenditure approaches, for the Euro area as a whole and for the four largest countries (Germany, France, Italy and Spain). The proposal is a large scale factor model such that the co-movements among the series are synthesized by a set of sixteen common factors, representing the Euro Area common trend, 4 country specific factors and 11 components specific factors. The statistical treatment is based on likelihood inferences for a suitable state space model that is able to accommodate temporal aggregation and any pattern of missing data at the end of the sample. The contribution is also innovative for the implementation of a computationally efficient algorithm able to address the increase of dimension of the state space form. In particular for the extraction of the common factors it is employed a reduction technique that decrease the dimension of the model to the number of common factors.

11.8.3 The application to the Euro Area

In this section we present the coincident index of the euro area economy resulting from an application based on a long monthly time series database provided by the Conference Board. Data are seasonal adjusted and concern:

1. employment from January 1987 to April 2014 in thousands of employees;

2. industrial production index from January 1987 to April 2014, base 2005 = 100;

3. retail sales series in volume from January 1995 to April 2014, base 2005 = 100;

4. index of manufacturing turnover from January 1995 to April 2014, base 2005 = 100.

In all the cases data are expressed in logarithms. Preliminary model selection suggested that a first-order autoregressive representation for the common factor and the idiosyncratic components is satisfactory. Even in presence of missing data at the beginning of the sample for retail sales and the index of manufacturing, the estimation period has been set to the full sample from January 1987 to April 2014.

The model shows good overall fit with parameters all significant. Values of model diagnostics are not shown here for reason of space, focusing to a comparative analysis of the SW-CEI with that of the Conference Board.

The upper part of Figure 11.1 shows the CEI obtained from estimation of the Stock and Watson methodology in comparison with that of the corresponding indicator of the Conference Board shown in the lower part of the same figure. Data in the graph are restricted to the period January 1995-April 2014. Main elements emerging from the comparison are the following:

1. despite different slopes among the two CEIs, turning points appear to be very close among the two indicators;

2. the Stock and Watson CEI is more erratic than that of the Conference Board;

3. the trough of the 2008-2009 recession is deepest for the Stock an Watson index. Same for the recovery. In general the amplitude of cycles are wider.

Then, we can conclude that the two approaches provide similar results even in presence of distinctive elements: in particular the evidence is that the modelling approach by Stock and Watson emphasizes the amplitude of cycles at price of a higher erraticity of the derived coincident indicator. By contrast according to the simpler and more direct application of the Conference Board the syntesis of common cycles among the component series smooth out more effectively their irregularities. Here the potential risk is of loosing some turning points with respect to the Stock and Watson coincident index even if it appears to be not the case of this application.

11.9 Conclusions

We have reviewed the main contributions of the Stock and Watson methodology for the construction of coincident, leading and recession composite indexes of the business cycle. We have understood that the most appealing feature of this theory is simply given by the formal definition of the target variable, which in the previous literature was not explicitly mentioned. We have discussed the complexity of the method, making use of a multivariate setup casted in a state space form and requiring the Kalman filter and smoothing technology for its estimation and inference. By contrast we have stressed how simple, clear and easy it is to interpret the classical procedure by The Conference Board for the construction of composite business cycle indicators. The discussion also presents the main elements of critique of the methods and reviews the main advancements. Our effort was to give the reader an opportunity to better understand the theory around the factor modelling approach for the construction of business cycle indicators, in the hope that who interested could be addressed towards a more independent and conscious analysis of pros and cons of the Stock and Watson approach.

We have also shortly reviewed the main applications of the Stock and Watson methodology outside the US. Finally, we have given space to the use of the Stock and Watson model in the applications of interest for statistical agencies and to a specific application to the euro area, comparing the derived coincedent index with that of the Conference Board. We have seen how applied results of this theory represent a powerful instrument for complementing standard official statistics with timely and indirect estimates of short term macroeconomic indicators. Therefore, these derived data represent an efficient way to answer to the continuous requests of statistics coming from economists, policy makers and the general public. The results of these studies are very encouraging and have shown the ability of the method to increase length, coverage and timeliness of short term statistics.

Bibliography

Akkoyun H. C., Gunay M. (2012) Nowcasting Turkish GDP Growth, Central Bank of the Republic of Turkey, working papers, 12/33, December 2012.

Altissimo F., Cristadoro R., Forni M., Veronese, G. (2010), New eurocoin: Tracking economic growth in real time. Review of Economics and Statistics, 92:1024-1034.

Burns A.F., Mitchell W.C. (1946), Measuring business cycles. NBER Studies in Business Cycles No. 2, New York.

Claus E., Claus I. (2007) Six Leading Indexes Of New Zealand Employment, CAMA Working Papers 2007-17, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Crone T. M., Clayton-Matthews A. (2004), Consistent Economic Indexes for the 50 States, Federal Reserve Bank of Philadelphia, Working Paper No. 04-9.

Dickey and Fuller (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association 74, no. 366, 427-431.

Diebold F.X., Rudebusch G.D. (1996), Measuring Business Cycle: A modern Perspective, The Review of Economic and Statistics 78 (February 1996), 67-77.

Durbin J., Koopman S.J. (2013), Time Series Analysis by State Space Methods, 2nd edition, New York: Oxford University Press.

Eklund B. (2007) Predicting recessions with leading indicators: An application on the Icelandic economy, Central Bank of Iceland, Economics Department, working papers, 33.

Engle and Granger (1987), Co-integration and error correction: Representation, Estimation and Testing, Econometrica 55, 251-276.

Fahrmeir L. (1992), Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models, Journal of the American Statististical Association, 97, 501-509.

Forni M., Hallin M., Lippi M., Reichlin L. (2000), The generalized factor model: Identification and estimation. The Review of Economics and Statistics 82 (4), 540-554.

Forni M., Hallin M., Lippi M., Reichlin L. (2001), Coincident and leading indicators for the Euro area, The Economic Journal 111 (May), C62-C85.

Forni M., Hallin M., Lippi M., Reichlin L. (2005), The generalized dynamic factor model: One-sided estimation and forecasting. Journal of the American Statistical Association, 100.

Frale C., Marcellino M., Mazzi G.L., Proietti T. (2010), Survey Data as Coincident or Leading Indicators, Journal of Forecasting, 29, 109-131.

Frale C., Marcellino M., Mazzi G.L., Proietti T. (2011), EUROMIND: a monthly indicator of the euro area economic conditions, Journal of the Royal Statistical Society, series A, 174, part 2, 439-470.

Fukuda S.I., Onodera T. (2001) A New Composite Index of Coincident Indicators in Japan: How Can We Improve the Forecast Performance?, CIRJE F-Series CIRJE-F-101, CIRJE, Faculty of Economics, University of Tokyo.

Gaudreault C., Lamy R., Liu Y. (2003) New Coincident, Leading and Recession Indexes for the Canadian Economy: An Application of the Stock and Watson Methodology, Department of Finance, working paper, 2003-12.

Gerlach S. and Yiu M. S. (2004) A Dynamic Factor Model for Current-Quarter Estimates of Economic Activity in Hong Kong, working papers 162004, Hong Kong Institute for Monetary Research.

Grassi S., Proietti T., Frale C., Marcellino M., Mazzi G.L. (2013), EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries, International Jounal of Forecasting, 2013.

Hall S. G., and Zonzilos N. G. (2003) An Indicator Measuring Underlying Economic Activity in Greece, Bank of Greece, working paper, no. 4.

Hamilton J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57, 357-384.

Harvey (1989), Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

Kim J.C., Nelson C.R. (1998), Business cycle turning points, a new coincident index and tests of duration dependence based on a dynamic factor model with regime switching, The Review of Economic and Statistics 80, 188-201.

Koopmans T.C. (1947), Measurement without theory. Review of Economics and Statistics 29, 161-179.

Marcellino M. (2006), Leading indicators, in Graham Elliott, Clive W.J. Granger and Allan Timmermann (Eds.), Handbook of Economic Forecasting, Volume 1, 2006 Elsevier B.V.

Mariano R.S., Murusawa Y. (2003), A new coincident index of business cycles based on monthly and quarterly series Journal of Applied Econometrics,18, 427-443.

Mitchell W., Burns A.F. (1938), Statistical indicators of cyclical revivals. NBER, New York. Reprinted in: Moore, G.H. (Ed.), Business Cycle Indicators. Princeton University Press, Princeton, 1961. Chapter 6.

Murphy A. P. (2005) An Economic Activity Index for Ireland: The Dynamic Single-Factor Method, Central Bank and Financial Services Authority of Ireland, Research Technical Paper, 4/RT/05, August.

Nieto F. H. (2003) Identifiability of a coincident index model for the Colombian economy, Borradores semanales de economïca, 194, Banco de la Republica, Bogota

Proietti T., Moauro F. (2006), Dynamic factor analysis with non-linear temporal aggregation constraints, Applied Statistics, 55, part 2, 281-300.

Reklaite A. (2011) Coincident, Leading and Recession Indexes for the Lithuanian Economy, Baltic Journal of Economics, 11(1), 91-107

Sims C. (1989), Comment to New indexes of coincident and leading indicators, in Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA, pp. 394-397.

Stock J.H., Watson M.W. (1988a), A probability model of the coincident economic indicators, NBER Working Papers Series No. 2772.

Stock J.H., Watson M.W. (1988b), Testing for Common Trends, Journal of the American Statistical Association, 83, No. 404, 1097-1107.

Stock J.H., Watson M.W. (1989), New indexes of coincident and leading indicators, in Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA, pp. 351-394.

Stock J.H., Watson M.W. (1991), A probability model of the coincident economic indicators, in Lahiri, K., Moore,G.H. (Eds.), Leading Economic Indicators: New Approaches and Forecasting Records. Cambridge University Press, Cambridge, UK, pp. 63-89.

Stock J.H., Watson M.W. (1992), A procedure for predicting recessions with leading indicators: econometric issues and recent experience, NBER Working Papers Series No. 4014.

Stock J.H., Watson M.W. (2006), Forecasting with many predictors, in Graham Elliott, Clive W.J. Granger and Allan Timmermann (Eds.), Handbook of Economic Forecasting, Volume 1, 2006 Elsevier B.V.

Zarnowitz V., Braun P. (1989), Comment to New indexes of coincident and leading indicators, in Blanchard, O., Fischer, S. (Eds.), NBER Macroeconomics Annual. MIT Press, Cambridge, MA, pp. 397-407.

IV | Turning Point Indicators

12

An Overview of Alternative Turning Points Composite Indicators

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ludovic Calès
European Commission, Joint Research Centre (JRC)
ludovic.cales@ec.europa.eu

Table of Contents

12.1 Introduction

The composite indicators presented in this part of the handbook aim to provide the probability to be in a given economic phase at a certain point in time. Consequently, based on the change from an economic phase to another, they allow to detect in real-time or to anticipate in a probabilistic way the occurrence of cyclical turning points. Thus this family of indicators are not intended for estimating a continuous variable mimicking the unobserved cyclical pattern of a given target variable, but rather identifying a discrete sequence of break points representing the change from a phase to another (turning points). In this context, turning points are viewed as discontinuities or break points within a given time series signalling the change from a phase (expansion or recovery) to another one (recession or slowdown) and vice versa. The mainstream turning points detection stems from the composite indicators presented in part three.

In such indicators, turning points are identified as relative maxima and minima of the composite indicators, in a given time-span, identified by means of a dating rule. The identification of the last relative maximum or minimum allows then to assess the occurrence of a new turning point, either for the present or the near future, as in De Bondt and Hahn (2014) for instance. This means that the way in which turning points are identified in the composite indicators presented in part 3 is very similar to the approach followed to derive historical turning points dating. In this part, we focus on rather different composite indicators allowing for turning point detection or forecasting. They provide a direct detection or forecasting of turning points, without passing by the estimation of the cyclical components as in part three.

The main advantage of these turning point composite indicators is to timely quantify the expectation of a change in the economic phase which indicates a turning point. Their main drawback is that they do not allow for a reliable estimation of the cyclical pattern between the turning points. By summarising the main difference between the indicators presented in part three and those presented in this part is represented by the following elements:

- The former provide a reliable estimate of the cyclical movement while indicators presented here do not.

- In the indicators presented in section 12.3, turning points are defined as relative maxima and minima of the composite indicators while here they are associated to a change of the regime in the composite indicators under the hypothesis that the dynamics of the composite indicators can be modelled as a regime-switching process.

The composite indicators presented here can complement those presented in the previous part, within a cyclical early warning system, even if inconsistencies between the two kinds of indicators can always occur. Such discrepancies could be eliminated by jointly modelling the two kinds of indicators developed in parts three and four within a complex non-linear model combining discrete and continuous variables. Nevertheless, since this is more a research field than a concrete methodology, we are not further considering it in this part.

The structure of the chapter is as follows:

In section 12.2 we are discussing the issues related to the identification of the reference variables as well as their relation to the construction of a reference turning points historical dating; section 12.3 is devoted to discuss the choice of the reference cycle(s) while section 12.6 discusses the main methodological approaches to be adopted when building up composite indicators for turning points detection; section 12.5 then presents a short literature review; finally section 12.6 provide a guide to this part and section 12.7 briefly concludes.

12.2 Identification of the reference variable

The identification of the reference variable for this kind of composite indicators is slightly less relevant than for those presented in the previous part. Nevertheless, it is useful to have a sufficiently clear idea of the

characteristic of the cycle for which we are trying to estimate the turning points. As already mentioned, the GDP could be the ideal candidate, however its unavailability at monthly frequency, at least in most cases, can make its utilisation quite complex. Indeed, detecting turning points on a quarterly basis would be inefficient and useless since policy makers and analysts require a higher frequency and more timely information.

Once again, the second best choice is the industrial production index even though the economic sentiment indicator could also be considered for the growth cycle (output gap) and the acceleration cycle. As mentioned in chapter 7 the availability of a monthly interpolated version of the quarterly GDP can also be used as a reference variable if timely and reliable enough.

The above considerations apply to the case where the goal is to build up a turning points detection composite indicator for the whole economy. In case where such indicators are applied to specific sectors or economic branches, the reference variable has to be modified accordingly with the characteristics of the sector/branch of interest. Similar considerations apply to the construction of composite turning points indicators for developing or under developed economies. As stated in chapter 7, for such kind of economies the reference variables can be really very country-specific reflecting the peculiarities of the economy under investigation.

Together with the identification of the reference variable, also the identification of the reference historical turning point dating is of crucial relevance for the construction of the composite indicators presented in this part. The availability of a reference historical dating of turning points for the cycle or the cycles of interest is essential to benchmark the composite indicators over a sufficient long time interval in order to assess its performance and to identify possible drawbacks (such as the tendency of providing false signals) which should be corrected, as much as possible, and of which users should be made clearly aware. In case where historical turning points dating have not been yet compiled, they should be computed as a prior step to the construction of turning points composite indicators (see Anas et al. (2017) (chapter 14) and chapter 14, annex for details. Historical dating can be computed either by using non-parametric dating rules such as in Bry and Boschan (1971) or its extension by Harding and Pagan (2002) or by parametric methods by using non-linear time-series models as proposed by Hamilton (1989). Further details on the construction of the historical datings are provided in Anas et al. (2017) (chapter 14), Anas et al. (2008) and Harding and Pagan (2011).

Here it is important to insist on a point which has been debated for several years. Both non-parametric and parametric dating rules have their advantages and drawbacks. In particular, since the composite turning points indicators are usually built up by using non-linear time-series models, the use of a parametric dating rule will ensure a better methodological consistency between the dating and the detecting exercises. On the other hand, by using a parametric rule, whenever new observations become available, the non-linear model used for the dating can be subject to revisions so that there is no guarantee that past turning points will be not revised. By contrast, the use of a non-parametric dating rule ensures that past turning points are never revised unless the input data are revised. Since one of the main theoretical requirements of an historical turning point dating is stability on the past, in our opinion, the use of a non-parametric dating rule can be preferable (see Anas et al. (2017) (chapter 14) and Anas et al. (2008)). A detailed presentation of parametric rules for turning points dating provided by Chauvet and Hamilton (2005) while a comparative evaluation of non-parametric and parametric dating rule can be found in Chauvet and Piger (2008).

Furthermore, interesting application of dating rules, with also a comparative investigation, to national economies are provided, as examples, see Anas et al. (2008) for the Euro area, Otsu (2017) for Japan, Krznar (2011) for Croatia, Bonenkamp et al. (2001) for the Netherlands, Kyung Sam and Lim Sung (2009) for South Korea and Tsouma (2010) for Greece.

Finally, we would like also to mention that an historical turning point dating is not necessarily an historical chronology. Even if it constitutes its core, an historical dating is mainly based on statistical evidence while a dating chronology needs an economic validation which is not necessarily relevant for the purpose of this part.

12.3 Identification of the reference cycle

Turning points composite indicators can be built up with reference to the classical business cycle and to the output gap (or growth cycle) as in the previous parts, but also referring to the acceleration cycle which was not considered before due to its high degree of volatility. The main difference with the indicators presented above is that turning points indicators can be constructed in order to detect or anticipate simultaneously classical business cycle and growth cycle turning points by exploiting the characteristic of the turning points sequence ABCD presented in Anas et al. (2017) (chapter 14) and introduced by Anas and Ferrara (2004). This possibility derives from the symmetric structure of the sequence of turning points.

By constructing composite turning point indicators for the three main reference cycles, we allow for a complete monitoring of the occurrence of cyclical turning points which constitutes an essential piece of a system of cyclical information. Furthermore, it is important to observe that, when considering the growth cycle as the reference cycle, the composite indicators constructed in this part are much less subject to the end-point estimate critics than those constructed in part three. Since here we do not necessary need to estimate the current and future evolution of the growth cycle but only its turning points, the impact of the de-trending methods is much less relevant. As shown in chapter 14, it is possible to build up turning points composite indicators of the growth cycle without making use at all of de-trending filters.

12.4 Methodology

Since we are aiming to construct composite indicators for identifying the probability of being, at present or in the near future, in a given economic phase and for estimating the rare events of turning points (change between phases), non-linear techniques appear the most suitable ones even if there are some contributions in the literature using linear methods. The use of non-linear techniques is also justified by the fact that cycles are often characterised by asymmetries across phases which cannot be properly detected when using linear modelling techniques. Within this class of models, the variable selection and the estimation are usually carried out in two consecutive steps. In order to properly select the most reliable variables, it is essential to have at disposal a previously computed turning points historical dating, so that variables can be ranked and selected according to their ability to reproduce historically the turning points sequence of the reference series. The variables selection methods considered are those previously presented in chapters 8 (U.S. Department of Commerce) and 10 (OECD) as well as in Anas et al. (2017) (chapter 14) and Mazzi (annex chapter 14). More generally, the framework for the variable selection see the one described in Guégan (2017) (chapter 5) and Kapetanios et al. (chapter 6).

Once the variables are selected, several modelling strategies can be used applied to construct the composite indicators. They can be grouped in the following categories:

- Binary regression models (Logit, PROBIT ...)

- Univariate and multivariate Markov-Switching models, also casted in a state space representation

- Thresholds time series models (TAR, SETAR)

- STAR models

A detailed review of non-linear modelling techniques for business cycle analysis is presented in Ferrara and Mazzi (2017) (chapter 13). The first three categories are the most widely used for constructing turning points composite indicators. If the individual component series are modelled separately, the composite indicators can be derived by pooling the results of each univariate model. The pooling, as in the previous part, can be based on a number of different weighting schemes and the results are quite sensitive to them. Alternatively, the composite turning point composite indicators can be obtained within a multivariate framework where the

selected variables are jointly modelled. This approach avoids the subjectivity related to the choice of the weighting scheme but it can be computationally more complex. When using non-linear models, it is important to carefully identify the most appropriate number of regimes to be used in order to better fit the cyclical characteristics of the reference variable.

Furthermore, it is also particularly useful to model in the most appropriate way the presence of the heteroskedasticity as well as to accurately define the switching mechanism. Since non-linear models can be quite complex to be estimated, in order to avoid convergence problems, especially for the estimation numerical algorithms, it is recommended to keep the models as parsimonious as possible by privileging, whenever possible, simple dynamic structure. When looking at several classes of non-linear models, the choice of the most appropriate model can be quite challenging. Billio et al. (2013) have compared the performance of Markov-switching models against SETAR models when detecting classical business cycle turning points. The results of such comparison have shown that, even if the SETAR model can outperform the Markov-switching ones in terms of detecting timeliness, they are much more subject to errors, especially type-1 error meaning the delivery of false turning points. For this reason, authors have considered that Markov-switching models should be preferred.

Finally, when using threshold models, the value of the threshold can play a very important role. In the large majority of models, especially when Markov-switching models are used, the natural threshold of 0.5 is adopted; see Anas et al. (2017) (chapter 14) and Anas et al. (2008). Moving the threshold in both directions can change the sensitivity of the model. Lowering the threshold can make the model more sensitive and increase the possibility of providing false signals. Increasing the threshold above 0.5 can reduce the sensitivity of the model and increase the possibility of missing turning points.

12.5 Review of the literature

The first attempts to probabilistically identify changes in the economic phases date from the late 80's with Hamilton (1989) and the Markov switching approach. Since then, a rich literature has flourished and the following briefly reviews the main approaches described in the previous section.

The first class of models are the binary regression models, often including a non-linear component. Taking into account breaks in logit models, Chauvet and Potter (2002), Chauvet and Potter (2005) and Chauvet and Potter (2010) could reproduce the US recession phases. The introduction of breaks greatly helps to identify the phases, indeed a single break already identifies the Great Moderation.[1] Dueker (2002) and Bellégo and Ferrara (2009) combine PROBIT and Markov switching models to forecast the US and euro area recessions, respectively.

These models evolved from a univariate form, see e.g. Dueker (1997), Estrella and Mishkin (1995) and Estrella and Mishkin (1998), to a multivariate form, see e.g. King et al. (2007), Bellégo and Ferrara (2012) and Chauvet and Potter (2010). The importance of financial leading variables, and in particular the term spread, in these models should be noticed.

The second class of models are the Markov switching (MS) models. Introduced in business cycle analysis by Hamilton (1989), these models have been widely used since then. These models offer the interesting advantage to identify automatically the different phases in time series without setting any threshold. By pooling the probabilities of recession stemming from univariate MS models, Anas and Ferrara (2004) propose a coincident indicator of US recessions later extended to the euro area in Anas et al. (2008). The main extensions on the Markov switching models are (i) time varying transition probabilities, see Filardo (1994) and Filardo and Gordon (1998) for factorial transition probabilities, see Layton and Smith (2007) and Bec et al. (2015) for transition probabilities based on functions, (ii) the dynamic factor Markov switching (DFMS) models which assume that a latent variable drives a set of variables and whose dynamics switch among

---

[1]See McConnell and Perez-Quiros (2000) for more details on the Great Moderation effect.

regimes, see Diebold and Rudebusch (1996). The second approach (DFMS) has gained some popularity and has been applied to the US and euro area, see Kim and Nelson (1998), Chauvet (1998), Chauvet and Piger (2008) and Camacho et al. (2012) among others.

The third class of models are the threshold models. The threshold autoregressive (TAR) models are piecewise linear processes whose parameters switch values when an observed transition variable passes a threshold. They have been introduced by Tong (1978). If the transition variable is the dependent variable, the model is called self-excited threshold autoregressive model (SETAR). While these models have been frequently used in business cycle analysis, few have been used to build turning points indicators. In business cycle analysis, we notice the works by Tiao and Tsay (1994) who use the TAR model to forecast the US GDP and highlight the recession and expansion phases, and by Proietti (1998) who uses a SETAR model to study the asymmetries of the business cycles in the US and compare it to a STAR model. The only turning points indicators based on SETAR models can be found in Ferrara and Guégan (2006) and Billio et al. (2013) which shows their ability in delivering timely signals.

The fourth class of models are the smooth transition autoregressive (STAR) models. Their data-generating process is a linear combination of autoregressive processes whose weights vary according to a continuous transition function summing to 1. The first turning point indicator developed with STAR models can be found in Terävirta and Anderson (1992) where each autoregressive process represent a phase of the business cycle. The method has then been widely used, see Stock and Watson (1999), and extended in order to take into account more non linearity and structural changes, see Lundbergh et al. (2003), Camacho (2004), Fok et al. (2005) among others.

12.6 Guide to the part

This section aims to introduce the part four of the handbook. Ferrara and Mazzi (2017) (chapter 13) review in details the methods presented in Section . An emphasis is put on the methodologies and in their applications in the development of turning points indicators. It also presents a review of the evaluations of the different models. The main classes of models are compared to the Markov switching approach. The general consensus is that the Markov switching models perform better than the other classes. Layton and Katsuura (2001) find that the binary response models provide poorer out-of-sample turning points forecasts, the binary response models being less flexible as requiring the complete knowledge of the reference cycle. Billio et al. (2013) find the SETAR models being fairly reactive but also very volatile and leading to false signals, while Deschamps (2008) finds that the MS are better to anticipate turning points.

An integrated approach for detecting in real-time turning points is then proposed by Anas et al. (2017) (chapter 14). After defining a an analytical framework as the sequence of turning points of business, growth and acceleration cycles, namely the $\alpha AB\beta CD$ approach, the authors consider both univariate and multivariate Markov switching models in the construction of turning point indicators for each of the three considered cycles. Regarding the univariate models, after a selection of the most informative variables for each cycle, they are modelled independently by using the best fitting Markov switching model.

The recession probabilities returned by each component series are then are pooled up, using a weighting scheme reflecting the relative reliability of each component series to obtain three composite indicators, one for each cycle. They provide independent sequences of turning points for the business (B and C turning points), growth (A and D turning points) and acceleration cycles (α and β). In the multivariate approach, the selected variables are jointly modelled by means of a multivariate Markov switching model in order to return a pair of composite indicators: one for the business and one for the growth cycle. Some criteria to assess the quality of the composite indicators are also identified and an empirical evaluation of the results obtained is provided.

Finally in the annex to chapter 14 a step by step approach to the construction of composite cyclical indi-

cators for turning points detection. For each step, after a description of the issue, some suggestions and recommendations are provided.

12.7 Conclusions

This chapter has introduced alternative ways to identify economic turning points as change of regimes, where each regime is associated to a specific cyclical phase. The approaches presented in this part are essentially probabilistic and not they likely require less subjective intervention than for composite indicators presented in the previous part. Their main advantage is to timely signal in real-time, or to anticipate, the significant changes in the cyclical economic situation. The main drawback of this approach is represented by the fact that they are not very much informative in-between turning points since they do not necessarily provide a reliable estimation of the cyclical pattern as it is the case for the indicators presented in part three. Nevertheless, especially MS models, can provide some information related to the size of the probability recession returned by the model, telling us if we are approaching the threshold or if we are far from it. Finally, it appears quite obvious that cyclical composite indicators presented in part 3 and turning points composite indicators presented in part 4, have to be intended as complementary tools in a cyclical early warning system.

Bibliography

Anas J., Billio M., Ferrara L. and Mazzi G.L. (2017), Chapter 14: Composite cyclical indicators detecting turning points within the ABCD framework, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Anas J., Ferrara L. (2004), Turning points detection: the ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, 1 (2), 1-36.

Anas J., Billio M., Ferrara L. and Mazzi G.L. (2008), A system for dating and detecting turning points in the euro area, The Manchester School, 76 (5), 549-577.

Bec F., Bouabdallah O. and Ferrara L. (2015), Comparing the shape of recoveries: France, the UK and the US, Economic Modelling, 44 (C), 327-334.

Bellégo C. and Ferrara L. (2009), Forecasting euro area recessions using time-varying binary response models for financial variables, Working Paper, Banque de France, No. 259.

Bellégo C. and Ferrara L. (2012), Macro-financial linkages and business cycles: A factor-augmented probit approach, Economic Modelling, 29, 1793-1797.

Billio M., Ferrara L., Guégan D. and Mazzi G.L (2013), Evaluation of Nonlinear time-series models for real-time business cycle analysis of the Euro area, Journal of Forecasting, 32 (7), 577-586.

Bonenkamp, J., Jacobs, J. and Kuper, G. H. (2001), Measuring business cycles in The Netherlands, 1815-1913: A comparison of business cycle dating methods, Working Paper, 2001, available online at: https://www.researchgate.net/profile/Gerard_Kuper/publication/4868341_Measuring_business_cycles_in_The_Netherlands_1815-1913_a_comparison_of_business_cycle_dating_methods/links/0fcfd510839a73565d000000.pdf.

Bry G. and Boschan C. (1971), Cyclical Analysis of Time Series: Selected Procedures and Computer Programs, New York, NBER.

De Bondt G. and Hahn E. (2014), Introducing the euro-Area Leading Indicator (ALI): real-time signals of turning points in the growth cycle from 2007 to 2011, Journal of Forecasting, 33, 47-68.

Camacho M. (2004), Vector smooth transition regression models for US GDP and the composite index of leading indicators, Journal of Forecasting, 23, 173-196.

Camacho M., Perez-Quiros G. and Poncela P. (2012), Markov-Switching dynamic factor models in real-time, Working Paper 1205, Banco de España.

Chauvet M. (1998), An economic characterization of business cycle dynamics with factor structure and regime switches, International Economic Review, 39, 969-996.

Chauvet M. and Hamilton J.D. (2005), Dating Business Cycle Turning Points, NBER Working Paper No. 11422.

Chauvet M. and Piger J. (2008), A comparison of the real-time performance of business cycle dating methods, Journal of Business Economics and Statistics, 26 (1), 42-49.

Chauvet M. and Potter S. (2002), Predicting recessions: Evidence of the yield curve in the presence of structural breaks, Economics Letters, 77 (2), 245-253.

Chauvet M. and Potter S. (2005), Forecasting recessions using the yield curve, Journal of Forecasting, 24 (2) 77-103.

Chauvet M. and Potter S. (2010), Monitoring business cycles with structural changes, International Journal of Forecasting, 26 (4) 777-793.

Deschamps P. (2008), Comparing smooth transition and Markov switching autoregressive models of US unemployment, Journal of Applied Econometrics, 23, 435-462.

Diebold F.X. and Rudebusch G.D. (1996), Measuring business cycles: A modern perspective, Review of Economics and Statistics, 78 (1), 67-77.

Dueker M. (1997), Strengthening the case for the yield curve as a predictor of US recessions, Federal Reserve Bank of St. Louis, Review, 79 (2), 41-51.

Dueker M. (2002), Regime-dependent recession forecasts and the 2001 recession, Federal Reserve Bank of St. Louis, Review, 29-36, November/December 2002.

Estrella A. and Mishkin F.S. (1995), The Term Structure of Interest Rates and Its Role in Monetary Policy for the European Central Bank, National Bureau of Economic Research Working Paper no. 5279, September.

Estrella A. and Mishkin F.S. (1998), Predicting US recessions: Financial variables as leading indicators, Review of Economics and Statistics, vol. 80, pp. 45-61.

Ferrara L. and Guégan D. (2006), Detection of industrial business cycles using SETAR models, Journal of Business Cycle Measurement and Analysis, 2 (3), 1-20.

Ferrara L. and Mazzi G.L. (2017). Chapter 13: Parametric Models for Cyclical Turning Point Indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Filardo A.J. (1994), Business cycle phases and their transitional dynamics, Journal of Business and Economic Statistics, 12, 299-308.

Filardo, A.J. and Gordon S.F. (1998), Business cycles duration, Journal of Econometrics, 85, 1, 99-123.

Fok D., van Dijk D. and Franses P.H. (2005), Forecasting Aggregates using Panels of Nonlinear Time Series, International Journal of Forecasting, 21(4), 785-794.

Guégan D. (2017), Chapter 5: Review of Parametric and non-Parametric Variable and Model Selection Techniques, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Hamilton J.D. (1989), A new approach to the economic analysis of non-stationary time series and the business cycle, Econometrica, 57, 357-384.

Harding D. and Pagan A. (2002), Dissecting the cycle: a methodological investigation. Journal of Monetary Economics, 49(2), pp. 365-381.

Harding D. and Pagan A.R. (2011), Econometric Analysis and Prediction of Recurrent Events, No 75, NCER Working Paper Series from National Centre for Econometric Research.

Kim C.J. and Nelson C.R. (1998), Business cycle turning points, a new coincident index, and tests of duration dependence based on dynamic factor model with regime-switching, The Review of Economics and Statistics, 80, 188-201.

King T.B., Levin A.T. and Perli R. (2007), Financial market perceptions of recession risk, FEDS Working Paper No. 2007-57.

Krznar, I. (2011), Identifying Recession and Expansion Periods in Croatia, Croatian National Bank, Working Papers W-29, Zagreb, November 2011.

Layton A. and Katsuura M. (2001), Comparison of regime switching, probit and logit models in dating and forecasting US business cycles, International Journal of Forecasting, 17, 403-417.

Layton A.P. and Smith D.R. (2007), Business cycle dynamics with duration dependence and leading indicators, Journal of Macroeconomics, 29, 855-875.

Lundbergh S., Teräsvirta T. and van Dijk D. (2003), Time-varying smooth transition autoregerssive models, Journal of Business and Economic Statistics, 21, 104-121.

Kyung Sam M. and Sung Joo L. (2009): Dating rules for turning points of growth cycles in Korea, OECD Journal: Journal of Business Cycle Measurement and Analysis, 1, available online at: http://dx.doi.org/10.1787/jbcma-v2009-art5-en.

McConnell M. and Perez-Quiros G. (2000), Output fluctuations in the United States: What has changed since the early 1980s, American Economic Review, 90,5, 1464-1476.

Otsu, T. (2017), Abridged Bry-Boschan Procedure, Seijo University economic papers (215), 85-126, 2017-01, available online at: http://www.seijo.ac.jp/education/faeco/academic-journals/jtmo420000001iji-att/215-otsu.pdf.

Proietti T. (1998), Characterizing asymmetries in business cycles using smooth-transition structural time series models, Studies in Nonlinear Dynamics and Econometrics, 3, 141-156.

Stock J. and Watson M. (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R. F. and White, H. (eds.), Cointegration, causality and forecasting. A festschrift in honour of Clive W. J. Granger. Oxford University Press: Oxford (UK).

Teräsvirta T. and Anderson H. (1992), Characterizing nonlinearities in business cycles using Smooth Transition Autoregressive Models, Journal of Applied Econometrics, 7, 119-136.

Tiao G.C. and Tsay R.S. (1994), Some advances in non-linear and adaptive modelling in time series, Journal of Forecasting, 13, 109-131.

Tong H. (1978), On a threshold model, in Pattern Recognition and Signal Processing, (ed.) C.H. Chen, Amsterdam: Sijhoff and Noordhoff.

Tsouma, E. (2010), The Greek Economy During 1970-2010 and the Recent Recession, Euroindicators Working Paper, presented at the 6th Eurostat Colloquium on Modern Tools for Business Cycle Analysis: The lessons from global economic crisis, held in Luxembourg, 26th-29th September 2010, available online at: http://ec.europa.eu/eurostat/documents/4187653/5782637/EWP-2011-032-EN.PDF/7ca28146-6b7b-4311-86ac-f0bdead93d52?version=1.0.

13

Parametric Models for Cyclical Turning Point Indicators

Authors

Laurent Ferrara
Banque de France
laurent.ferrara@banque-france.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Table of Contents

13.1 Introduction

Generally, practitioners in business cycle analysis assume that economic cycles can be seen as an alternation of two conjonctural phases, namely a phase of high economic activity and a phase of low economic activity. Those economic cycles can be the business, growth or acceleration cycles, as defined in Anas et al. (2008) or in Anas et al. (chapter 14). For example, in the case of business cycles as defined in the NBER sense, the high phase corresponds to expansions while the phase of low economic activity corresponds to recessions. Some papers in the literature (Sichel (1994) or Ferrara, 2008) argue that there may be more than two phases in the cycle, but we assume in this paper that two phases are sufficient to provide a realistic description and that extensions to more than two regimes are straightforward in subsequent models. In addition, we assume here that we observe the values of a binary variable (r_t) that takes value 1 when the economy belongs to the low phase of the cycle at date t and 0 otherwise.

Generally, the chosen dates for the turning point chronology are those stemming from reference institutions, such as the NBER for the US and CEPR for the euro area. International dating chronologies are also provided by the Conference Board, OECD and the Economic Cycle Research Institute (ECRI). Thus, a peak in the cycle is located at a given date t_0 if $r_{t_0} = 0$ and $r_{t_0+1} = 1$, while a trough in the cycle is located at date t_1 if $r_{t_1} = 1$ and $r_{t_1+1} = 0$.

The objective of parametric models in business cycle analysis is to provide, at each date t, an estimated probability of being in a specific phase of the cycle at $t + h$, denoted $\hat{P}_{t+h}$ for any horizon $h \geq 0$, based on a given set of information. In turn this probability can be used to develop a turning point indicator. If $h = 0$, this corresponds to a nowcasting exercise and the indicator is coincident with the cycle, otherwise this probability can be seen as a leading indicator of a turning point in the cycle.

In this respect, new tools have been put forward in business cycle analysis, mainly based on non-linear parametric modelling. Non-linear models have the great advantage of being flexible enough to take into account some specific stylized facts of the economic business cycle, such as asymmetries in the phases of the cycle. In this respect, emphasis has been placed on the class of non-linear dynamic models that accommodate the possibility of regime changes. Especially, Markov-Switching (MS) models have been popularized by Hamilton (1989) and have experienced a great success due to their ability to reproduce accurately business cycle phases. Besides the well-known Markov-switching approach, another parametric model allows for different regimes in business cycle analysis, the threshold autoregressive (TAR) model, introduced by Tong (1990) or the smooth transition autoregressive (STAR) model, put forward by Teräsvirta (1994), which can describe the asymmetry observed in economic variables (see also Tiao and Tsay (1994), Proietti (1998), or Granger et al. (2010)).

Such models differ from MS models in the sense that the variable governing changes in regimes is observed, leading thus to easier statistical inference. In addition to those previous models, binary response models, that relate a quantitative variable to a binary one, are often considered to replicate economic cycles. In this chapter, we present in detail those four types of models in separate sections. In each section we first describe the models by taking into consideration the simple specification that allows for only two regimes, knowing that it can be easily generalised to more regimes. Second, we consider statistical inference by assuming that all the processes are covariance-stationary. Then we present the latest extensions put forward in the literature, especially by focusing on multivariate extensions. Last we propose some applications of the considered models for business cycle analysis and the construction of cyclical turning points. A last section contains a qualitative assessment of the alternative models for cyclical turning point indicators. Note that further details of the models presented in this chapter can be found in the books of Tong (1990), Hamilton (1994), Wooldridge (1994), Franses and van Dijk (2000) or Granger et al. (2010).

13.2 Binary response models

Binary response models, such as logit and probit models, are well known statistical models and are widely used in business cycle analysis. Those models aim at establishing a relationship between one explanatory variable (x_t) and the binary dependent variable (r_t) that describes the occurrence of a given phase of the cycle. As an output the model delivers a probability of being in one of the two states according to the values of the explanatory variable.

13.2.1 The standard model

An important aspect of binary response models in business cycle analysis is that the phases of the cycle have to be known *a priori*, before the analysis. Generally, an exogenous dating chronology of turning points is used, stemming from official institutes (NBER, ECRI, among others). Thus we assume that we observe the values of a binary variable, denoted (r_t), that takes value 1 when the economy belongs to a low phase of the cycle at date t and 0 otherwise. Logit and probit models rely on the assumption that the values of the binary dependent variable, (r_t) stems from a latent continuous variable, denoted (y_t), defined by the following linear equation:

$$y_t = \beta_0 + \beta(B)x_t + \varepsilon_t, \tag{13.1}$$

where $\beta(.)$ is a p-order polynomial such that $\beta(B) = \beta_1 I + \beta_2 B + \ldots + \beta_{p+1}B^p$, B being the standard backward operator and $p \geq 0$ being the autoregressive lag that controls the persistence, where for all t, (x_t) is an explanatory stationary variable and where (ε_t) is the error term supposed to be a strong white noise process with finite variance σ_ε^2. The distribution of (ε_t) is discussed below. The observed binary variable (r_t) is supposed to be linked to the latent variable (y_t) by the following relationship:

$$r_t = \begin{cases} 1 & \text{if } y_t \leq 0, \\ 0 & \text{if } y_t > 0 \end{cases} \tag{13.2}$$

At each t, it can be easily proved that the probability that a low phase of the cycle occurs, conditionally to the observations of (x_t), is given by:

$$P(r_t = 1 | x_t, x_{t-1}, \ldots) = F(-\beta_0 - \beta(B)x_t), \tag{13.3}$$

where $F(.)$ is the cumulative density function of the variable (ε_t). For example, the probit model is defined by assuming that the error term (ε_t) is Gaussian, that is $F(.) = \phi(.)$ the cumulative density function (cdf) of the standard Gaussian distribution. The shape of the function $F(.)$ allows to discriminate between various binary response models. For example, the well known logit model uses for $F(.)$ the logistic function defined for any real z by :

$$F(z) = \frac{\exp(z)}{1 + \exp(z)}. \tag{13.4}$$

Both logistic and cdf standard Gaussian functions allow to plug the quantitative information contained in (x_t) into the interval $[0, 1]$. The logistic function appears to be smoother than the cdf standard Gaussian in the sense that the transition from 0 to 1 takes more time. In other words, the cdf standard Gaussian is closer to the indicator function $1_{(z>0)}$ describing a discrete transition from 0 to 1.

13.2.2 Inference for standard models

Regarding parameter estimation for binary response models, we refer, for instance, to the monography of Wooldridge (1994). We assume that we observe the time series $(x_1, \ldots, x_T)$. The general log-likelihood

function for the logit/probit model described by equations (13.1)-(13.3) is given by the following equation:

$$L(\theta) = \sum_{t=1}^{T} r_t \times \log(P(r_t = 1|x_t)) + (1 - r_t) \times \log(P(r_t = 0|x_t)) \tag{13.5}$$

where $\theta = (\beta_0, \beta_1, \ldots, \beta_p, \sigma_\varepsilon^2)$ and where T is the finite sample size of the information set on which the statistical inference is based.

Generally, parameter estimation of linear dichotomous models is carried out by the maximum likelihood estimation (MLE) method. Thus, the ML estimate of the (p+2)-dimensional parameter θ is the value such that:

$$\hat{\theta}_{ML} = \arg\max L(\theta) \tag{13.6}$$

where $L(.)$ is the log-likelihood given by equation (13.5). From a technical point of view, a solution of this equation is obtained by using an iterative maximisation algorithm such as the Newton-Raphson algorithm. Under the independence assumption and some other classical assumptions, it can be proved that the ML estimate is asymptotically consistent and Normally distributed. In the framework of more complex binary response models, Chauvet and Potter (2002) propose the use of the Gibbs sampling methodology in order to evaluate the log-likelihood.

To assess the goodness-of-fit of binary response models, some statistical measures are also available. For instance, information criteria allow a comparison of different models. Some well known criteria are the Akaike information criteria (AIC) and the Schwartz information criteria (SIC), respectively defined by:

$$AIC = -2L(\hat{\theta}) + 2k, \tag{13.7}$$

$$SIC = -2L(\hat{\theta}) + k\log(n). \tag{13.8}$$

Estrella (1998) also proposes a $Pseudo - R^2$ measure given by the following expression:

$$Pseudo - R^2 = 1 - \frac{L_u}{L_c}^{-(2/T)L_c} \tag{13.9}$$

where L_u is the log-likelihood of the considered model and L_c is the log-likelihood of a reference nested model to which L_u is compared. By construction, the nested model must have a lower log-likelihood value than the basic model. This $Pseudo - R^2$ measure is often used in the literature for comparison between models.

13.2.3 Extensions of binary response models for business cycle analysis

Many extensions of binary response models have been proposed in the statistical literature to account for some observed stylized facts. Without being exhaustive, we present some of them that have proved useful for business cycle analysis.

Richer dynamics

First, some researchers pointed out the importance of allowing for a dynamic serial correlation in the model, in order to improve prediction recession performances (see Dueker (1997), or Moneta (2005)). As in Dueker (1997), the dynamic structure of the model can be extended by adding a lagged dependent variable to the classical model (13.1)-(13.3). Thus, the dynamic dichotomous model takes the following general form:

$$P(r_t = 1|x_t) = F(-\beta_0 - \beta(B)x_t - \delta(B)r_t), \tag{13.10}$$

where δ is the following polynomial of degree q: $\delta(B) = \delta_1 B + \ldots + \delta_q B^q$. However, in practice, the polynomial δ is generally of the form $\delta(B) = \delta_p B^q$, q being empirically chosen according to the data.

Using this model, Dueker (1997) points out the predictive power of the slope of the bond yield curve in comparison with four other variables: the Commerce Department index of leading indicators, real M2 growth, the spread between the 6-month commercial paper and 6-month treasury bill rates and the Standard and Poor's 500 index of stock prices. Dueker (1997) confirms the result of Estrella and Mishkin (1995) by showing that the slope of the yield curve is the dominant predictor of recessions at horizons beyond three months. Estrella and Mishkin (1998) point out that, even if many financial variables result in a relatively good in-sample fit, in most cases a parsimonious probit model with the interest rate spread as the sole predictor tends to produce the most accurate out-of-sample prediction of US recessions.

Duarte et al. (2005) estimate probabilities of recession in the euro area from 1970 to 2000 by using both US and euro term spread. They interestingly conclude that the euro term spread dominates the US spread with parameter estimates being more than twice the respective estimates of the US coefficients. Moreover, they note that being in recession one year ago increases the probability of being in expansion at the current date. This result is due to the fact that recessions have a duration lower than one year. Note however that the inclusion of the lagged dependent variable (r_t) into the model tends to render every other parameters associated with different covariates non-significant. This is due to the strong auto-correlation of the dependent variable within each phase of the cycle.

Recently, Kauppi and Saikkonen (2008) have suggested a way to extend the classical specification based on the idea that the impact of the explanatory variable (x_t) may depend on a lagged value of the binary variable (r_t). For example, they point out that the impact of the interest rate spread may be asymmetric and differ during recession and expansion periods. The specification proposed by the authors that accommodates this extension is the following:

$$P(r_t = 1 | x_t) = F(-\beta_0 - \beta(B)x_t - \delta(B)r_t - \gamma r_{t-d}x_t), \tag{13.11}$$

where $\gamma r_{t-d}x_t$ is the interaction term that occurs for a delay d, γ measuring the interaction effect. Parameter estimation of the model is carried out with classical ML method for sake of simplicity. An application of this model can be found in Nyberg (2010) in order to predict US and German probability of recession, conditionally on information conveyed by stock prices as well as domestic and foreign term spread. In addition, the author finds that the spread between long term interest rates in Germany and the US can be a useful predictor for German recessions.

To account for autocorrelation effects in the binary-dependent models, Dueker (1997), Dueker (2002) comes up with the idea to allow parameters to evolve through time according to a Markov-Switching model. Let's assume the existence of an unobserved variable (s_t) that follows a Markov chain taking the values 0 or 1 such that the transition probabilities are given by:

$$P(s_t = 0 | s_{t-1} = 0) = p, P(s_t = 1 | s_{t-1} = 1) = q. \tag{13.12}$$

Thus,

$$r_t = \begin{cases} 1 & \text{if } \varepsilon_t \leq -\beta_0(s_t) - \beta(s_t)'x_t, \\ 0 & \text{if } \varepsilon_t > -\beta_0(s_t) - \beta(s_t)'x_t \end{cases} \tag{13.13}$$

The Conference Board's composite index of leading indicators (CLI) is a leading index composed of the macroeconomic variables that have been proven to anticipate the US business cycle, such as consumers expectations, stock prices or building permits Dueker (2002) uses the CLI to forecast the US recessions in 1991 and 2001 through a probit model whose parameters evolve according to a Markov chain with two states. Dueker (2002) points out two regimes of low and high variance. He shows that the 2001 recession could have

been predicted using this kind of models. Another application of this time-varying binary response model can be found in Bellégo and Ferrara (2009) on euro area data.

In a series of papers, Chauvet and Potter (2002), Chauvet and Potter (2005), Chauvet and Potter (2010) have shown the usefulness of the introduction of breaks into a logit model in order to reproduce the US recession phases. In Chauvet and Potter (2002), Chauvet and Potter (2005) they use the term spread variable to anticipate the US recessions by including one or several breaks, while in Chauvet and Potter (2010) they use the usual four series considered by the NBER Dating Committee, namely IPI, employment, real personal income and real manufacturing and trade sales. The introduction of a single break into the model takes into account the *Great Moderation* effect, a decrease in the amplitude of the cycles pointed out in several papers (see for example McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004)).[1] The date of the break is located in 1984Q3 and an endogeneous detection procedure identifies a break that occurred between 1977 and 1982. The model is specified as follows for equation (13.1):

$$y_t = \begin{cases} \beta_0^1 + \beta^{1'} x_t + \varepsilon_t^1 & \text{if } t \leq t_0 \\ \beta_0^2 + \beta^{2'} x_t + \varepsilon_t^2 & \text{if } t > t_0 \end{cases} \tag{13.14}$$

where t_0 is the date of the break and (ε_t^1) and (ε_t^2) are Gaussian white noises with $\sigma_{\varepsilon^1}^2 \geq \sigma_{\varepsilon^2}^2$. The introduction of multiple breaks into the model via equation (13.1) allows to consider the changes in the characteristics of each cycle overtime.

Multivariate framework

We assume now that we observe a vector y_t of k explanatory variables, $y_t = (x_t^1, \ldots, x_t^k)'$. In the multivariate framework, we get the following equation in place of (13.1):

$$\begin{aligned} y_t = & \beta_0 + \alpha_1^1 x_t^1 + \ldots + \alpha_p^1 x_{t-p}^1 + \ldots \\ & + \alpha_1^k x_t^k + \ldots + \alpha_p^k x_{t-p}^k + \varepsilon_t, \end{aligned} \tag{13.15}$$

where $(\varepsilon_t)_t$ is the error term supposed to be a strong white noise process with finite variance σ_ε^2. King et al. (2007) prove evidence that bivariate ($k = 2$) probit models that include both term and credit spreads outperform other models in recession prediction (with an horizon of 1 year). Indeed, they note that the simultaneous occurrence of a flat or inverted yield curve and a high credit spread has been a strong signal of an imminent recession. Filardo (1999) uses a 4-dimensional vector of explanatory variables containing term spread, corporate spread, S&P500 return and growth of the CLI, while Chauvet and Potter (2010) consider the four variables analysed by the NBER Dating Committee (IPI, employment, income and sales). Chauvet and Potter (2010) show that it is not possible to reproduce the NBER recession phase of the US economy by simply putting those variables in a multivariate logit model. They have to strongly constraint the logit model to be able to reproduce the reference dating chronology over the past, by taking several breaks into account. This constraint implies the use of a complicated estimation procedure based on Gibbs sampling.

When the number of available explanatory time series is large, a possibility is to aggregate probabilities stemming from univariate models to get an unique synthetic probability of recession. Bellégo and Ferrara (2009) show that such an approach can be fruitfully applied to forecast euro area recessions. King et al. (2007) also conclude that a bayesian averaging of recession probabilities improve the accuracy of prediction in comparison with simple averaging. An alternative to combined forecasts is to implement a factor-probit model in order to reduce the dimension of the problem through a dynamic factor (see for example, Bellégo and Ferrara (2009), Bellégo and Ferrara (2012), or Chen et al. (2011)). The objective of dynamic factor modelling is to decompose the k-vector (x_t) into a sum of two mutually orthogonal unobservable components:

---

[1] The *Great Moderation* effect relates in fact to a substantial reduction in macroeconomic volatility as documented for example by McConnell and Perez-Quiros (2000). By extension, this also refers in the empirical literature to a decrease in the amplitude of the cycles, but the connection is not clearly stated.

a common component of low dimension (χ_t), summarizing the dynamics common to all the series, and an idiosyncratic component (ξ_t), specific to each series. The common component $(\chi_t)_t$ is supposed to linearly summarize the common behavior of the k series. For $t = 1, \ldots, T$, the static factor model is defined by

$$x_t = \Lambda f_t + \xi_t, \tag{13.16}$$

where Λ is the loading matrix of dimension $(n \times r)$, the common component $\chi_t = \Lambda f_t$ is driven by a small number r of factors f_t common to all the variables in the model such that $f_t = (f_{1t}, \ldots f_{rt})$, and $\xi_t = (\xi_{1t}, \ldots, \xi_{kt})'$ is a vector of k idiosyncratic mutually uncorrelated components, driven by variable-specific shocks. We refer to Stock and Watson (2002) or to Forni et al. (2005) for estimation issues. Thus, (13.1) can be rewritten as:

$$y_t = \alpha + \beta_0' f_t + \ldots + \beta_k' f_{t-k} + \varepsilon_t, \tag{13.17}$$

where for $j = 0, \ldots, k$, $f_{t-j} = (f_{1,t-j}, \ldots, f_{r,t-j})'$ is a r-vector of lagged factors used as explanatory variables, $\beta_j = (\beta_j^1, \ldots, \beta_j^r)'$ is a r-vector parameter and $(\varepsilon_t)_t$ is the error term supposed to be a strong white noise process with finite variance σ_ε^2.

13.2.4 Applications of standard binary response models

Many applications in business cycle analysis have been carried out by using standard univariate binary response models. Especially, several empirical studies have shown evidence that the interest rates term structure is a very powerful tool in predicting business cycle turning points (we refer for example to Stock and Watson (2003), Estrella et al. (2003), Rudebusch and Williams (2009), and the references therein). Generally, the variable considered is either the yield curve or the spread between long-term interest rates (generally 10 years) and short-term interest rates (generally 3 months) The yield curve plots the yield of treasury bonds against their maturity and is therefore typically upward sloping and somewhat convex. However, in the case of recession, the yield curve tends to be flat or slope downward.

This idea of using term spreads to predict recessions was largely exploited for example in the papers of Estrella and Hardouvelis (1991), Estrella and Mishkin (1995), Estrella and Mishkin (1998), Bernard and Gerlach (1996),Bernard and Gerlach (1998), Dueker (1997), Attah-Mensah and Tkacz (1998), Moneta (2005), Chauvet and Potter (2005), Kauppi and Saikkonen (2008) or Rudebusch and Williams (2009). By using a univariate binary-dependent variable model all these authors proved the empirical evidence that the term structure of interest rates can be useful to detect recessions with a significant lead, up to eight quarters in advance.

These kinds of studies are carried out for many different countries. Attah-Mensah and Tkacz (1998) focus on Canada, Dueker (1997), Estrella and Mishkin (1998), Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) focus on the United States while Estrella and Mishkin (1995) and Bernard and Gerlach (1996),Bernard and Gerlach (1998) consider many countries including Japan, Canada, the United States and some European countries. Estrella et al. (2003) compare Germany and the US and show that before 2000 recession prediction models are stable in both countries. Concerning the Euro area as a whole, Moneta (2005) has provided evidence that the term spread was a good predictor for recession in the zone, by using both a standard probit model and the extension that includes lagged dependent variable. Moneta (2005) also proves that considering not only aggregated Euro area data but also individual countries could improve the forecasts. The fact that the euro area has not been widely considered in empirical studies is mainly due to the lack of long historical data for the euro area as whole.

Many other variables, mainly financial ones, have been considered in the empirical literature as leading indicators of recession. For example, by using 54 different financial-market variables, including all of the series considered by Stock and Watson (2003) as well as a variety of corporate credit spreads and several other distinct measures of market liquidity, King et al. (2007) show that corporate credit spreads are at least as informative as term spreads to predict US recessions over the past two decades. Nyberg (2010) also consider

stock prices and foreign term spread for US and Germany, while Bellégo and Ferrara (2009) consider a set of 13 variables for the euro area as a whole.

13.3 Markov-Switching models

The covariance-stationary Markov-Switching (MS) model has been first introduced by Quandt (1958), then reconsidered by Neftçi (1982) or Neftçi (1984) and popularized in economics by Hamilton (1989). The success of the Markov-Switching model, as underlined by Hamilton (1989) or Hamilton (1990), relies on its ability to reproduce the US business cycle instead of using a complex and non-transparent procedure as implemented by the NBER Business Cycle Dating Committee. A detailed description of MS models can be found in Hamilton (1994) or Krolzig (1997). Below we present the model, discuss statistical inference issues, give useful extensions for business cycle analysis then comment applications of MS models to the development of cyclical indicators.

13.3.1 The Markov-Switching model

The univariate process (x_t) follows a two-regime Markov-switching model, denoted MS(2) - AR(p), if it verifies the following equation:

$$x_t - \mu(S_t) = \sum_{i=1}^{p} \phi_i(S_t)(x_{t-i} - \mu(S_{t-i})) + \sigma(S_t)\varepsilon_t, \tag{13.1}$$

where the non-observed process (S_t) is an ergodic Markov chain and where (ε_t) is a standardized white noise process. The parameters $\mu(S_t)$, $\phi_1(S_t)$, ..., $\phi_p(S_t)$ and $\sigma(S_t)$ are thus time-varying and describe the dependence of the process (x_t) to the current regime S_t. S_t characterizes the unobserved state, or regime, of the economy at date t. Each regime corresponds to a given phase of the economic cycle and for each date t, the economy can either stay in the same regime at date $t+1$ or switch to the other regime. The associated transition matrix of the Markov chain (S_t) is defined by:

$$P[S_t = j | S_{t-1} = i] = p_{ij}, \tag{13.2}$$

with $0 < p_{ij} < 1$ for $i, j = 1, 2$ and $\sum_{j=1}^{2} p_{ij} = 1$. Transition probabilities are collected in the transition matrix η given by:

$$\eta = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} \tag{13.3}$$

In the two-regime case, the unconditional probabilities associated to the process (x_t) are equal to:

$$P[S_t = 1] = \frac{1 - p_{22}}{1 - p_{11} + 1 - p_{22}} = \pi, \tag{13.4}$$

and

$$P[S_t = 2] = 1 - \pi. \tag{13.5}$$

The estimation step enables to get for each date t the forecasted probability, the filtered probability and the smoothed probability of being in a given regime i, respectively given by $P(S_t = i | \hat{\theta}, x_{t-1}, \ldots, x_1)$, $P(S_t = i | \hat{\theta}, x_t, \ldots, x_1)$ and $P(S_t = i | \hat{\theta}, x_T, \ldots, x_1)$, where $\hat{\theta}$ is the estimated vector of parameters of the model.

13.3.2 Statistical inference

Starting from a finite sample $(x_1, \ldots, x_T)$, testing for Markov-Switching type of non-linearity is quite challenging. Many statistical tests have been put forward in order to test of linearity against a MS alternative and/or to assess the number of regimes (see e.g. Hansen (1996), Hamilton (1996), or Garcia (1998)). Under the null of linearity, the issue is that some parameters are not identified, such as the transition probabilities, leading thus to non-standard asymptotic distributions. Recently, Carrasco et al. (2009) have proposed a class of optimal tests for the constancy of parameters in random coefficients models. Their testing procedure covers the class of Hamilton models, where the parameters vary according to an unobservable Markov chain. An important advantage of their SupTS test is that it only requires estimating the model under the null hypothesis where the parameters are constant. Empirical critical values can be computed from a large number of iterations (10 000 for example) for a sample size equal to the size of the original data set. The values of the parameters used to simulate the series are those obtained from the estimation of the model under the linear null.

Regarding the parameter estimation issue, the maximum likelihood method is used in connection with the Expectation Maximization (EM) algorithm. The EM algorithm is an iterative technique for maximizing the likelihood function in case of models with missing observations or models where the observed time series depends on some unobservable latent stochastic variables. Hamilton (1990) reports that the EM algorithm is, in general, relatively robust with respect to poorly chosen starting values of the parameters, quickly moving to a reasonable region of the likelihood surface. Furthermore, as shown in Hamilton (1990), the EM algorithm can be used in conjunction with the filtering and smoothing algorithm to draw inference over the state allocation under the simplifying assumption of knowing the parameter of the Markov-switching model. In practice, it turns out that convergence issues often occur when estimating MS models. This is generally due to a ill-behaved likelihood function that can present several local maxima with higher likelihood than the MLE. For example, Boldin (1996) points out this phenomena. In addition, he underlines the lack of robustness to the sample of the MS model originally used by Hamilton (1989) for the US GDP.

Parameter estimation of the model (13.1), namely $\mu(S_t)$, autoregressive coefficients $\phi_i(S_t)$, variances $\sigma(S_t)$ and transition probabilities p_{ij}, is carried out by maximum likelihood estimation (MLE) applied to the conditional density:

$$f(x_t/F_{t-1}, \theta) = \sum_{j=1}^{2} f(x_t/S_t = j, F_{t-1}, \theta) P(S_t = j/F_{t-1}, \theta), \tag{13.6}$$

where θ is the vector of parameters, F_{t-1} is the information set until date $t-1$, $F_{t-1} = (x_1, \ldots, x_{t-1})$ and $P(S_t = j/F_{t-1}, \theta)$ is referred to as the filtered probability of being in state j. The issue relies on the fact that the variable S_t cannot be observed and has thus to be estimated for each date t. In this respect, Hamilton (1994) suggests to implement a recursive algorithm allowing the estimation of filtered and smoothed probabilities for each date t. Assuming that the initial value $P(S_1 = i/F_0)$ and parameter θ are known, Hamilton (1994) shows that optimal inference and forecasting of the filtered probability can be respectively obtained by iterating the two following equations:

$$P(S_t = i/F_t, \theta) = \frac{P(S_t = i/F_{t-1}, \theta) \cdot f(x_t/S_t = i, F_{t-1}, \theta)}{\sum_{j=1}^{2} P(S_t = j/F_{t-1}, \theta) \cdot f(x_t/S_t = j, F_{t-1}, \theta)} \tag{13.7}$$

and

$$P_{t+1/t} = \eta P_{t/t} \tag{13.8}$$

where $P_{t/t} = (P(S_t = 1/F_t, \theta), P(S_t = 2/F_t, \theta))'$ where η is the transition matrix given in (13.3).

The specification of the model can be carried out by maximizing a given criterion, like AIC, the log-likelihood

or by minimizing the p-value of a test that rejects either the null hypothesis of linearity of the data or the non-correlation of the residuals.

13.3.3 Extensions

Various extensions of Markov-switching processes have been proposed in the literature, ranging from time-varying transition probabilities (see Filardo and Gordon (1994), Diebold et al. (1994), Layton and Smith (2007)) to multivariate models. We do not present here all the extensions, but we focus on most used extensions in business cycle analysis.

A first basic idea is to recognize that transition probabilities p_{ij}, $i, j = 1, 2$, are not necessarily constant over-time and that it may be useful to allow them to evolve through time. This comes from the fact the probability of exiting of a recession increase with time. Empirical evidence of such result was put forward by Sichel (1991) through a hazard function approach, but he shows that while recessions tend to exhibit the duration dependence property, the expansion phases do not. In the MS framework, time-varying (TV) transition probabilities can be broadly defined as follows:

$$p_{ijt} = P[S_t = j | S_{t-1} = i, z_{t-1}], \tag{13.9}$$

where z_t is a k-vector of exogenous variables, that can be duration variables or leading indicators. A general way to express the TV probabilities, for $i, j = 1, 2$, is the following:

$$p_{ijt} = \frac{1}{1 + \exp(-\beta_{i0} - \sum_{s=1}^{k} \beta_{is} z_{s,t-1})}, \tag{13.10}$$

where β is $(k + 1)$-vector of coefficients. The logistic function used here ensure that the probability stays between zero and one. This approach is considered by Filardo (1994), Durland and McCurdy (1994), Diebold et al. (1994), Filardo and Gordon (1998) or Layton and Smith (2007). Against this background, Durland and McCurdy (1994) investigate the issue of duration dependence in the US business cycle and show that recessions are duration dependent whilst expansions are not. Layton and Smith (2007) develop a Markov-Switching model whose transition probabilities evolve through time according to a logit functional form. The logit model for the transition probabilities include the duration as explanatory variables as well as leading indicator of the US cycle, namely the Short Leading Index (SLI) and the Long Leading Index (LLI) computed by the ECRI.

Bec et al. (2015) propose a MS model to account for the various shapes of recoveries at the end of a recession. It relies on the bounce-back effects first analyzed by Kim et al. (2005) and extends the methodology by proposing *i)* a more flexible bounce-back model, *ii)* explicit tests to select the appropriate bounce-back function, if any, and *iii)* a suitable measure of the permanent impact of recessions. In the U-shaped bounce-back model, denoted BBU, the equation for $\mu(S_t)$ becomes:

$$\mu(S_t) = \gamma_0 + \gamma_1 S_t + \lambda \sum_{j=1}^{m} \gamma_1 S_{t-j}, \tag{13.11}$$

where the m and λ parameters govern respectively the duration and the magnitude of the bounce-back effect. For the V-shaped recession model, denoted BBV, the bounce-back function takes the form:

$$\mu(S_t) = \gamma_0 + \gamma_1 S_t + (1 - S_t)\lambda \sum_{j=1}^{m} \gamma_1 S_{t-j}. \tag{13.12}$$

Finally, the expression of μ_t in the "depth" bounce-back model, denoted BBD, is:

$$\mu(S_t) = \gamma_0 + \gamma_1 S_t + \lambda \sum_{j=1}^{m}(\gamma_1 + x_{t-j})S_{t-j}. \tag{13.13}$$

The value of the bounce-back parameter, λ, is crucial for the shape of the recovery. First, it is worth noticing that all these models differ from the standard model if and only if $\lambda \neq 0$. Then, for a bounce-back effect to occur, this parameter must be negative: in this case, the last term of the right hand side of the three equations above is positive and makes the growth rate larger for the quarters immediately following a recession. Bec et al. (2014) use this model to evaluate bounce-back effects in France, US and UK using real GDP series.

Suppose now that we observe a k-vector y_t of stationary variables, k being relatively small. Krolzig (1997) put forward an extension of the equation (13.1) in the multivariate framework. We define the k-dimensional second order process (y_t) as a MS(2)-VAR(p) model if it verifies the following equation:

$$y_t - \bar{\mu}(S_t) = \sum_{i=1}^{p} \Phi_i(S_t)(y_{t-i} - \bar{\mu}(S_{t-i})) + \epsilon_t, \tag{13.14}$$

where $(\epsilon_t)_{t \in Z}$ is a multivariate white noise Gaussian process with variance-covariance matrix $\Sigma(S_t)$, where $\bar{\mu}(S_t)$ is a k-vector of conditional means depending on S_t and where $\Phi_1(S_t), \ldots, \Phi_p(S_t)$ are $k \times k$ matrices describing the dependence of the model to the regime S_t. The full representation of the model requires the specification of the variable $(S_t)_t$ as a first order Markov chain with two regimes. This multivariate approach has been implemented by Krolzig (2003) to construct a turning point chronology for the euro area as whole and by Ferrara (2003) to develop an indicator of the US business and growth cycles.

Alternatively, the information contained in the k variables can be summarized into an univariate underlying factor, supposed to represent the common evolution of all the series, which switches between two distinct regimes according to a Markov chain. This Dynamic Factor Markov-Switching (DFMS) model was first sketched by Diebold and Rudebusch (1996), while theoretical and empirical aspects are widely discussed in Kim and Nelson (1998). For a single common factor, we define the model as follows, for $n = 1, \ldots, k$:

$$x_t^n = \gamma_n f_t + u_t^n, \tag{13.15}$$

with

$$\phi(B)f_t = \mu(S_t) + \varepsilon_t, \tag{13.16}$$

where γ_n are referred to as the loadings, $(u_t^n)_t$ is supposed to follow a Gaussian stationary AR(1) process with finite variance σ_n^2, (ε_t) is a Gaussian white noise process with unit variance and $\phi(B) = I - \phi_1 B - \ldots - \phi_p B^p$. We assume that $(f_t)_t$ and the idiosyncratic noises $(u_t^n)_{n=1,\ldots,k}$ are non-correlated and that the idiosyncratic noises $(u_t^n)_{n=1,\ldots,k}$ are not cross-correlated. Parameter estimation of this model can be carried out either simultaneously, as proposed by Kim and Nelson (1998), or in two steps, by estimating first the common factor $(\hat{f}_t)_t$ and then by fitting a univariate MS(2)-AR(p) process on the estimated factor. This latter approach is generally used in empirical papers (see Darné and Ferrara (2011)) due to its simplicity. Indeed non-convergence estimation issues often arise during the simultaneous estimation procedure, even if the number k of variables is low.

13.3.4 Applications

Many applications have been carried out using MS models in business cycle analysis, starting from the seminal work of Hamilton (1989). This work received a great attention because Hamilton showed that a quite simple econometric model was able to reproduce the US business cycle chronology estimated by the NBER Dating Committee through a complex process involving expert claims. In this respect, many papers have tried to construct a turning point chronology based on MS models (see for example, Krolzig (2003), Billio et

al. (2008) or Anas et al. (2008)). Especially, as an output of the MS model, we get the estimated filtered probability of being in a given regime. Thus the estimated probability of being in the low phase of the cycle is given by:

$$\hat{P}(r_t = 1) = P(S_t = 1|\hat{\theta}, x_t, \ldots, x_1), \tag{13.17}$$

where $S_t = 1$ is identified as the low regime of growth and $S_t = 2$ as the high regime. Based on this probability, inference can drawn by economists about the current state of the economy in real time. Due to the huge number of papers dealing with business cycle analysis based on MS models, it is quite impossible to carry out an exhaustive review of applications in macroeconomics and finance. Anyway, some empirical works specifically focus on the construction of turning point indicators and their evaluation.

As regards the broadest measure of economic activity, namely GDP, Chauvet and Hamilton (2006) use a MS filter in real-time to assess in real-time each quarter the probability of US recessions. Based on this single variable, they show that this approach nicely replicates the US business cycle and is robust in real-time. They put forward a rule to interpret in real-time the probability to avoid false signals and propose to wait until the next quarter to decide whether the current quarter is in recession or not. In spite of very good results, they note however that this system introduces a lag and is quite coincident with NBER announcements, leading thus to a limited interest for short-term economic analysts. Billio and Casarin (2010) and Billio and Casarin (2011) also consider MS model in order to identify in real time business cycle turning point in the euro area. Nalewaik (2012) tries to adopt another measure of the whole economic activity based on incomes, referred to as Gross Domestic Income (GDI). He showed that for the latest 2008-09 US recession a MS model applied to GDI performs better than applied to GDP but evidence is less clear-cut for past periods.

Assume now we observe a k-vector y_t of stationary variables, well known by practitioners to describe the cycle. If k is relatively small, let's say $k \leq 6$, a MS-VAR model as described by equation (13.14) can be estimated to detect turning points. For example, Darné and Ferrara (2011) put forward a coincident indicator of the euro area acceleration cycle by estimating a VAR model using the industrial confidence indexes of the six main euro area countries (Germany, France, Italy, Spain, Netherlands and Belgium). Also, a DFMS model given by equations (13.15) and (13.16) can be implemented to deal with several time series. Many CCIs have been developed based on this type model. For example, in their seminal paper, Kim and Nelson (1998) reproduce the US business cycle using a small dataset of of four series (industrial production, employment, retail sales and incomes). Such series have been also used by Chauvet (1998) and Chauvet and Piger (2008) to build CCIs based on DFMS models. The advantage of DFMS models is that they account for a larger number of series. For example Camacho et al. (2014) estimate the real-time probability that the euro area enters a recession based on 13 variables, sampled at both quarterly and monthly frequencies (see also Camacho et al. (2012), for DFMS models that take into account mixed frequencies and ragged-edge data issues). The real-time experience of DFMS models is described in Hamilton (2011) who points out that the decision rule adopted by the authors (a recession is recognized as soon as the probability exceeds 80% during three consecutive months) lead in real-time to a lagged signal. When dealing with k variables, Billio et al. (2012) propose a different approach based on a combination scheme to predict turning points in the bayeasian framework.

Alternatively, Anas et al. (2008) have proposed a set of two CCIs for the euro area for both the business (BCCI) and growth (GCCI) cycles. Both indicators are constructed starting from the k probabilities of turning points estimated from k univariate MS models applied to each variable entering the composite indicator (see details in the next chapter of this volume). A similar approach has been implemented by Anas and Ferrara (2004) to build a coincident indicator of US recessions.

13.4 Threshold models

In this chapter we focus on the threshold autoregressive (TAR) model introduced at the beginning of the eighties by Tong (1978) and Tong and Lim (1980). The TAR model is a piecewise linear process whose main

characteristic is its ability to switch the parameters of a linear autoregressive model according to the value of an observable transition variable. A detailed presentation and discussion of such models can be found in the monography of Tong (1990).

13.4.1 The standard model

The univariate process (x_t) follows a two-regime threshold autoregressive process, denoted TAR(2,p), if it verifies the following equation:

$$
\begin{aligned}
x_t = {} & (1 - I(z_{t-d} > c))(\phi_{1,0} + \sum_{i=1}^{p} \phi_{1,i} x_{t-i} + \sigma_1 \varepsilon_t) \\
& + I(z_{t-d} > c)(\phi_{2,0} + \sum_{i=1}^{p} \phi_{2,i} x_{t-i} + \sigma_2 \varepsilon_t),
\end{aligned}
\tag{13.1}
$$

where c is the threshold, $d > 0$ the delay, (ε_t) a standardised white noise process and (z_t) the transition variable that can be observed. Here, $I(.)$ is the indicator function such that $I(z_{t-d} > c) = 1$ if $z_{t-d} > c$ and zero otherwise. If, $\forall t$, $z_t = x_t$, the process is referred to as self-exciting TAR process (SETAR). For a given threshold c and the position of the random variable z_{t-d} with respect to this threshold c, (x_t) follows here a particular AR(p) model. The model parameters are $\phi_{j,i}$, for $i = 0, \ldots, p$ and $j = 1, 2$, the standard errors σ_1 and σ_2, the threshold c and the delay d.

Using some algebraic notations, the model (13.1) can be rewritten as a regression model. Denote $I_d(c) \equiv I(z_{t-d} > c)$, $\Phi_1 = (\phi_{1,0}, \cdots, \phi_{1,p})'$, $\Phi_2 = (\phi_{2,0}, \cdots, \phi_{2,p})'$ and $\mathbf{x'_{t-1}} = (1, x_{t-1}, \cdots, x_{t-p})$, then, we get the following alternative representation from (13.1):

$$
x_t = (1 - I_d(c))\mathbf{x'_{t-1}}\Phi_1 + I_d(c)\mathbf{x'_{t-1}}\Phi_2 + ((1 - I_d(c))\sigma_1 + I_d(c)\sigma_2)\varepsilon_t.
\tag{13.2}
$$

A major difficulty in applying TAR models is the specification of the transition variable (z_t), which plays a key role in the non-linear structure of the model. Since there is only a finite number of choices for the parameters c and d, the best choice can be done using an information criterion such as the Akaike Information Criterion (AIC).

13.4.2 Statistical inference

We assume now that we observe $(x_1, \cdots, x_T)$ data stemming from equation (13.2) with $\sigma_1 = \sigma_2 = \sigma$. It is generally useful to test evidence of non-linearity in the data against the linear benchmark alternative. To retain a SETAR$(2, p)$ model against an AR(1) model, we can test evidence of two regimes against a single one. Several testing procedures have been put forward in the literature, we refer for example to Hansen (1999) or Granger et al. (2010) for a review. The relevant null assumption is:

$$
H_0 : \phi_{1,i} = \phi_{2,i}, \forall i.
$$

It is well known that this testing problem is complicated because nuisance parameters, especially the threshold c, are not identified under the null. This leads to non-standard asymptotic distributions for the tests, which in turn implies that simulation-based are necessary to get correct inference. For example, Tsay (1989) suggests the use of ordered regressions according to the values of the transition variable. The Tsay test is based on the recursive residuals and involves a standard F-statistic to test the null hypothesis of linearity. Hansen

(1997) considers another approach based on a Likelihood Ratio test (LR test). If the errors are i.i.d., a test with near-optimal power against alternatives distant from the null hypothesis is the standard F-statistic:

$$F_T = T(\frac{\tilde{\sigma}_T^2 - \hat{\sigma}_T^2}{\hat{\sigma}_T^2}),$$

where

$$\tilde{\sigma}_T = \frac{1}{T} \sum_{t=1}^{T} (x_t - \mathbf{x}'_{\mathbf{t-1}} \hat{\Phi}_1),$$

where $\hat{\Phi}_1$ is the least squares (LS) estimate for the parameter Φ_1 under H_0. Since F_T is a monotonic function of $\hat{\sigma}_T^2$, it is easy to see that

$$F_T = sup_{c \in C} F_T(c),$$

where

$$F_T(c) = T(\frac{\tilde{\sigma}_T^2 - \hat{\sigma}_T^2(c)}{\hat{\sigma}_T^2(c)}),$$

is the pointwise F-statistic against the alternative hypothesis $H_1 : \phi_{1,i} \neq \phi_{2,i}$, when $c \in C$ is known. Since c is not identified, the asymptotic distribution of F_T is not χ^2 and derives from Hansen (1996). The corresponding critical values cannot in general be tabulated since this distribution depends on unknown moments functionals.

The testing procedure is non-standard as c (and also d) are nuisance parameters that are unidentified under the null hypothesis. When the assumed values of c and d are far from their true values, under the alternative, the test lacks power. Then, we can calculate the statistics over a grid of values for the nuisance parameters (c, d) and take the supremum or some average of the statistic F_T over a predefined set C of admissible values. Boostrap simulation methods can also be used to get the values of F_T for the chosen grid of values for the nuisance parameters (c, d). The LR test permitting to test existence or not of threshold is described in Hansen (1997) and the way to compute the distribution law using boostrap methods can be found in this latter paper or in Hansen and Seo (2002). Note also that the CUSUM test adapted by Tsay (1989) enables to detect non-linearity with threshold and that a Lagrange Multiplier test has been also proposed by Proietti (1998). Last, as discussed in Tsay (1989), some scatterplots are often useful to provide information about both the number of regimes and the threshold values.

The equation (13.2) is a regression equation (albeit non-linear in parameters) and an appropriate estimation method for the parameters is the LS method. Denote that under the assumption that (ε_t) is a Gaussian strong white noise, LS is equivalent to the maximum likelihood estimation. Since the regression equation (13.2) is non-linear and discontinuous, the easiest method to obtain the LS estimates is to use sequential conditional LS.

In the following we assume for simplicity $\sigma_1^2 = \sigma_2^2$. For a given value of c, using the notation $\mathbf{x}'_{\mathbf{t-1}}(c) = [\mathbf{x}'_{\mathbf{t-1}}(1 - I_d(c)), \mathbf{x}'_{\mathbf{t-1}}(I_d(c))]$ and $\Phi = (\Phi_1, \Phi_2)$, the LS estimates for the parameters Φ are :

$$\hat{\Phi}(c) = \left(\sum_{t=1}^{T} \mathbf{x}'_{\mathbf{t-1}}(c) \mathbf{x}_{\mathbf{t-1}}(c)^{-1} \right) \left(\sum_{t=1}^{T} \mathbf{x}'_{\mathbf{t-1}}(c) x_{t-1}(c) \right), \tag{13.3}$$

with residuals $\hat{\varepsilon}_t(c) = x_t - \mathbf{x}'_{\mathbf{t}}(c) \hat{\Phi}(c)$, and the residual variance is:

$$\hat{\sigma}_T^2(c) = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t(c)^2. \tag{13.4}$$

The LS estimate of c is the value that minimizes equation (13.4):

$$\hat{c} = \arg\min_{c \in C} \hat{\sigma}_T^2(c),$$

(13.5)

where $C = [C_1, C_2]$, C_1 and C_2 are real numbers.

Under these assumptions, for given d and threshold value c, the LS estimates $\hat{\Phi}(c)$ converge to $\Phi(c)$, a.s., as $n \to \infty$. It can be proven, under other regular conditions, that the LS previous estimates are $n^{-1/2}$ asymptotically consistent (see Tsay (1989), Chan (1993) and Hansen (1997)). In this latter case, the estimated threshold parameter is consistent and tends to the true value at rate T and suitably normalized follows asymptotically a Compound Poisson process. Tests can be used to improve the properties of the threshold estimates in order to build robust confidence intervals. Using Monte Carlo methods, Enders et al. (2007) provide confidence intervals for the threshold parameter. Kapetanios (2000) investigates also the performance of the conditional least square estimator for small sample TAR models. He shows that the threshold parameter exhibits large biases, especially when the constant terms in the model are different from zero.

In practice, we first need to determine the parameters c, d, p_1, p_2 in order to estimate all the parameters of the model (13.1). We can proceed in the following way:

- We assume that P is the maximum possible order of the two sub-regimes and D is the greatest possible delay.

- The threshold parameter c is chosen by grid-search procedure. The grid points are obtained using the quantiles of the sample under investigation. One uses generally equally spaced quantiles from the 10 (percent) quantiles and ending at the 90 (percent) quantiles.

- Now, for each fixed pair $(d, c_i), 0 < d < D, i = 1, \cdots m$, the appropriate TAR model has to be identified.

- The AIC criterion is used for selection of the orders p_1 and p_2.

  In this context, the used AIC criterion is given by:

  $$AIC(p_1, p_2, d, c) = \ln\left(\frac{1}{n}\sum_{t=1}^{n}\hat{\epsilon}_t^2\right) + 2\frac{p_1 + p_2 + 2}{n},$$

  (13.6)

  where $\hat{\epsilon}_t$ denotes the residuals.

Finally the model with the parameters p_1^*, p_2^*, d^* and c^* that minimize the AIC criterion can be retained. Since for different d there are different numbers of values that can be used for estimation, the following adjustment should be done, with $n_d = \max(d, P)$, $P = \max(p_1*, p_2*)$:

$$AIC(p_1^*, p_2^*, d^*, c^*) = \min_{p_1, p_2, d, c} \frac{1}{n - n_d} AIC(p_1, p_2, d, c).$$

(13.7)

In principle, this procedure can be used to specify the number of regimes, except that the grid-search procedure is very computationally demanding when the number of regimes is strictly greater than two. In this respect, Gonzalo and Pitarakis (2002) propose a procedure in which the number of regimes are selected sequentially.

13.4.3 Extensions

An obvious extension of SETAR models consists in allowing more than two regimes in the model. This may be useful for example when one aims at taking specific periods of time into account. Also, to add more flexibility, a

specific AR(p) model can be assumed in each separate regime, with p depending on the regime. Sometimes, the transition variable z_t is a transformation of the dependent variable, such as for example, $z_t = x_t - x_{t-1}$. Thus, in this latter case, the variable in differences acts as transition variable.

Bec et al. (2014) put forward a threshold model that includes a bounce-back function, analogous to the one for MS models presented in equations (13.11), (13.12) and (13.13), in order to replicate high growth rates observed just after recessions. They use this bounce-back SETAR model to predict some European GDP just after the 2008-09 recession and they show that accuracy can be increase by comparison with a linear benchmark model.

Obviously, the univariate TAR model can be generalised to the multivariate TAR model that account for several variables. Suppose now that we observe a k-vector y_t of stationary variables, k being relatively small. Tsay (1998) defines the general multivariate TAR model with r regimes as follows:

$$y_t = \sum_{i=1}^{r}\{\mu_i + \sum_{j=1}^{p}\mathbf{\Phi_{ij}}y_{t-j} + \varepsilon_{it}\}I(c_{i-1} < z_t < c_i), \tag{13.8}$$

where $(c_i)_i$ are the thresholds, $(\mu_i)_i$ are intercept vectors of dimension k, $\mathbf{\Phi_{ij}}$ are $k \times k$ parameter matrices.

13.4.4 Applications

With regards to the application of threshold models in business cycles analysis, we can quote the work of Tiao and Tsay (1994) who use SETAR models to provide out-of-sample multi-step-ahead forecasts for the quarterly US real GDP from 1947 to 1990. They highlight the different behavior of US GDP during recession and expansion periods. This work is the first one which uses this class of model to investigate the business cycle. Potter (1995) and Hansen (1997) also describe the US GDP with a TAR model. Enders et al. (2007) study the behavior of real US GDP growth and discuss existence of persistence growth regimes based on confidence intervals build around the threshold parameter. Hansen (1997) and Montgomery et al. (1998) explore the non-linearities in the business cycle through the use of a SETAR model for US monthly unemployment rate. Proietti (1998) detects asymmetry inside the business cycle with SETAR models using quarterly US industrial production series from 1947 - 1996. Hansen (1999) study US monthly industrial production for the period 1960-1989 using SETAR models with two and three regimes.

In spite of many applications in business cycle analysis, few papers consider in fact the use of threshold models to develop turning point indicators. In this latter framework, the estimated probability of being in the low phase of the economic cycle is given by:

$$\hat{P}(r_t = 1) = I(z_{t-d} \leq \hat{c}). \tag{13.9}$$

This probability only takes as values 0 or 1, as no smooth transition between the two regimes is allowed. This can be seen as a drawback of this approach. However, statistical inference is easy in the sense that it is based on LS estimation, the transition variable being observed. Ferrara and Gueguan (2005) apply a SETAR model to the euro area industrial production index in order to detect the low phase of the industrial business cycle referred to as the industrial recessions. They show that this approach leads to fruitful results in terms of business cycle replication.

Billio et al. (2013) carry out a horse-race between SETAR and MS-AR models to monitor business cycles in the euro area using the IPI and the unemployment rate. They implement a real-time comparison exercise and they show that, in spite of higher volatility in the emitted signal, threshold models are useful to detect turning points in business cycles. Especially, they send a timely signal in real-time.

13.5 Smooth transition models

In the previous section we have introduced threshold models enabling to account for a transition between regimes using a discrete transition mechanism. However, it turns out that sometimes a smooth transition may appear more realistic and more flexible. In this respect, smooth transition autoregressive (STAR) models have been put forward and described by Teräsvirta and Anderson (1992), Teräsvirta (1994), Proietti (1998). A review of recent developments has been done by van Dijk et al. (2002) and a very complete description of such models can be found in Granger et al. (2010).

13.5.1 The standard model

The process (x_t) follows a general STAR process if its autoregressive data-generating mechanism is such that:

$$x_t = \Phi_1 \mathbf{x}'_{\mathbf{t-1}}(1 - G(z_t, \gamma, c)) + \Phi_2 \mathbf{x}'_{\mathbf{t-1}} G(z_t, \gamma, c) + \varepsilon_t, \tag{13.1}$$

using the notations Φ_1, Φ_2 and $\mathbf{x}'_{\mathbf{t-1}}$ introduced in the previous section. The variable (z_t) is any observable variable considered as the transition variable. The function $G(z_t, \gamma, c)$ denotes a continuous transition function and (ε_t) represents the error term. We assume here the same variance in each regime, but this assumption can be easily relaxed. Note that if $z_t = x_{t-d}$ and G is the indicator function, then we get the SETAR process introduced in equation (13.2).

A very popular choice for the transition function is given by the logistic function such that:

$$G(z_t, \gamma, c) = (1 + exp(-\gamma(z_t - c)))^{-1}, \quad \gamma > 0. \tag{13.2}$$

With increasing values for z_t, the function (13.2) changes monotonically from 0 to 1 with the so-called threshold parameter c locating a well balanced situation between both regimes as $G(z_t = c, \gamma, c) = 0.5$. The parameter γ can be interpreted as the steepness parameter which determines the speed of transition between both regimes. For $\gamma \to \infty$, the function (13.2) collapses into a discrete indicator function whereas equation (13.1) reproduces the TAR model given in equation (13.2). Another odd monotically increasing function from 0 to 1 is the standard Normal cumulated distribution function $N((z_t - c)/\sigma)$, where σ is a scale parameter. Note that any cumulated distribution function could be used.

In case of an even function, we can also consider the exponential function such that:

$$G(z_t, \gamma, c) = 1 - (exp(-\gamma(z_t - c)^2)). \tag{13.3}$$

For small or large values of z_t, G is close to 1. When $\gamma \to \infty$, then $G \to 1 - I(z_t = c)$. This function is generally used to account for three regimes in the data; the low and high regimes (e.g. recessions and fast expansions) having the same dynamics while the intermediate regime (e.g. moderate expansions) possesses its own dynamics.

In this chapter, we focus on the smooth transition autoregressive (STAR) model described by:

$$\begin{aligned} x_t = &(\phi_{1,0} + \sum_{i=1}^{p} \phi_{1,i} x_{t-i})(1 - G(x_{t-d}, \gamma, c)) \\ &+ (\phi_{2,0} + \sum_{i=1}^{p} \phi_{2,i} y_{t-i}) G(x_{t-d}, \gamma, c) + \varepsilon_t. \end{aligned} \tag{13.4}$$

Equations 13.1 and 13.2 corresponds to the "Logistic Smooth Transition Autoregressive" (LSTAR) model, and equations 13.3 and 13.4 corresponds to the "Exponential Smooth Transition Autoregressive" (ESTAR) model.

13.5.2 Statistical inference

In this section we discuss estimation and testing procedures which are generally nested to obtain parameter estimates for the model (13.1).

Testing procedures for STAR model

Concerning the testing procedures, we refer to Luukkonen, Saikkonen and Teräsvirta (1988), Teräsvirta (1994) and Proietti (1998). Testing the linearity against a SETAR specification is discussed in the previous chapter. We have seen that it is a highly nonstandard inferential problem. More convenient approximate test procedures are available for STAR alternatives. When the irregular variance is constant, the LM test of linearity against LSTAR or ESTAR alternatives is a test of $\gamma = 0$ in the model:

$$x_t = \phi_{1,0} + \phi_{2,0}G(x_{t-d}) + \sum_{j=1}^{p}(\phi_{1,j} + \phi_{2,j}G(x_{t-d}))x_{t-j} + \varepsilon_t, \tag{13.5}$$

assuming $\varepsilon_t \sim N(0, \sigma^2)$. This testing procedure is also non-standard, since under the null the parameters $\phi_{2,0}$, $\phi_{2,j}$ and c are not identified. It is noteworthy that the block of the information matrix corresponding to these parameters is null, which violates the standard regularity conditions under which the LR test is derived.

The way of getting around this problem consists of first deriving the LM statistic as a function of the unidentified parameters $LM(\phi_{2,0}, \phi_{2,j}, c)$ over all possible values $(\phi_{2,0}, \phi_{2,j}, c)$. When the transition variable is known, this yields the test statistic TR^2, where R^2 is the coefficient of determination in the regression of the LS residuals $\hat{\varepsilon}_t = x_t - \hat{\phi}_{1,0} - \sum_{i=1}^{p}\hat{\phi}_{1,j}x_{t-j}$, on $(1, x_{t-1}, \cdots, x_{t-p}, x_{t-1}x_{t-d}, \cdots, x_{t-p}x_{t-d})$. The coefficients associated to the cross-product terms depend on $\phi_{2,j}$ and c, but not on $\phi_{2,0}$. Hence, under the null the test has a $\chi^2(p)$ distribution, but the fact that $\phi_{2,0}$ does not enter the coefficients is responsible for the low power of the test when the non-linearity is mainly due to the intercept (i.e. $\phi_{2,0}$ is large and the $\phi_{2,j}$ values are small).

A general way for circumventing the lack of identifiability while enhancing the power properties of the test against the alternative amounts to replacing $G(x_{t-d}, \gamma, c)$ by a first-order Taylor approximation around $\gamma = 0$.

For the LSTAR model, when the transition variable is known, the LM test takes the usual TR^2 form, where R^2 is the coefficient of determination in the regression:

$$x_t = \phi_{1,0} + \sum_j \alpha_{1,j}x_{t-j} + \sum_j \alpha_{2,j}x_{t-j}x_{t-d} + \sum_j \alpha_{3,j}x_{t-j}x_{t-d}^2 + \sum_j \alpha_{4,j}x_{t-j}x_{t-d}^3 + \varepsilon_t. \tag{13.6}$$

The null here is : $H_0 : \alpha_{2,j} = \alpha_{3,j} = \alpha_{4,j} = 0, j = 1, \cdots p$. The test statistic has an asymptotic $\chi^2(3p)$ distribution. Alternatively the modified LM test ,

$$F_T = \frac{(SSE_0 - SSE_1)/3p}{SSE_1/(T - 4p - 1)},$$

can be used, since it yields an F-test statistic with better size properties. SSE_0 denotes the sum of squared estimated error from the full regression (13.6) and SSE_1 is the sum of the square residual when we regress y_t on the past and the products.

Nested tests are developed to discriminate between LSTAR and ESTAR models in Eitrheim, O., Terasvirta, T. (1996).

Parameters estimation for STAR model

In order to estimate the parameters of the model (13.1), we need to specify the transition function $G(y_{t-d}, \gamma, c)$ and to choose an appropriate threshold c. Generally the modelling process starts with a test parameter constancy, such as testing whether a STAR model would be more appropriate than a simple linear AR model.

Assuming for instance that a LSTAR model with an autoregressive order $p = 1$ and two regimes is the preferred model, a test for parameter constancy under assumptions H_0 and H_1 is given by:

$$H_{LSTAR,0} : \phi_{1,0} = \phi_{2,0}, \quad \phi_{1,1} = \phi_{2,1}.$$

Parameters of the transition function, γ and c, are not involved in the null hypothesis, yielding thus unidentified nuisance parameters.

Now, consider the null hypothesis of linearity as a test of $H_{AR,0} : \gamma = 0$, in which $G(y_{t-d}, 0, c) = 1/2$, so that the STAR model (13.1) can be written as:

$$x_t = \frac{1}{2}(\phi_{1,0} + \phi_{2,0}) + \frac{1}{2}(\phi_{1,1} + \phi_{2,1})x_{t-1} + \varepsilon_t, \tag{13.7}$$

which is linear, regardless of the truth of $H_{LSTAR,0}$. Thus, to be sure of the null we want to consider, it is important to include parameters in the transition function for purposes of testing. This last problem can be avoided by expressing the transition function by a Taylor expansion around $\gamma = 0$, which is a simple but nevertheless a little bit technique operation.

When the transition function and the threshold variable have been determined, the parameters in STAR models can be estimated by Non-Linear Least Squares (NLLS) method.

We assume that we observe $(x_1, \cdots, x_T)$ and we rewrite equation 13.4 as:

$$x_t = H(\mathbf{x_{t-1}}, \Theta) + \varepsilon_t.$$

Thus the NLLS estimator is given by:

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{t=1}^{T}(x_t - H(\mathbf{x_{t-1}}, \Theta))^2 = \arg\min_{\Theta} \sum_{t=1}^{T} \varepsilon_t^2,$$

where $\mathbf{x_t} = (1, x_{t-1}, \cdots, x_{t-p})$. If the innovations $(\varepsilon_t)_t$ are Gaussian, then the NLLS estimator is equivalent to the quasi-maximum likelihood estimator. These former estimators are consistent and asymptotically Normal under appropriate regularly conditions.

13.5.3 Extensions

In the literature, extensions of smooth transition models are fewer than those of MS models. A particular case of model (13.1) is obtained for $z_t = t/T$, where T is the number of observations. This yields a linear model with deterministically changing parameters. Such a model has a role to play, among other things, in testing

parameter constancy in linear models. A univariate model of this type will be called the time-varying STAR model by Lundbergh et al. (2003) and a special case when $\gamma \to \infty$ is a linear model with breaks.

A possible extension is to extend the number of non-linear components to get a multiple regime STAR model (MRSTAR) as proposed by van Dijk and Franses (1999) such as:

$$
\begin{aligned}
x_t =& \Phi_0 \mathbf{x_{t-1}}' + \Phi_1 \mathbf{x_{t-1}}' G(z_{1t}, \gamma_1, c_1) + \Phi_2 \mathbf{x_{t-1}}' G(z_{2t}, \gamma_2, c_2) \\
&+ \Phi_{12} \mathbf{x_{t-1}}' G(z_{1t}, \gamma_1, c_1) G(z_{2t}, \gamma_2, c_2) + \varepsilon_t,
\end{aligned}
\tag{13.8}
$$

This model takes a third regime into account, that can be interpreted as a moderate regime of growth, in addition to the two extreme regimes, that can be interpreted as recession and rapid expansion regimes. An interesting special case of the MRSTAR model (13.8) is obtained for $z_{2t} = t$. This model is called the time-varying STAR (TV-STAR) model for example by Lundbergh et al. (2003). This specific model nests both the standard STAR model (when $\gamma_2 = 0$) and the time-varying AR model (when $\gamma_1 = 0$).

As for the TAR model, the univariate STAR model can be generalised to the multivariate STAR model. In this respect, a logistic vector STAR (LVSTAR) model of dimension k can be defined as follows:

$$
y_t = \mu_0 + \sum_{i=1}^{p} \mathbf{\Phi_j} y_{t-j} + \mathbf{G}(\mathbf{z_t}, \gamma, \mathbf{c})(\mu_1 + \sum_{i=1}^{p} \mathbf{\Psi_j} y_{t-j}) + \varepsilon_t,
\tag{13.9}
$$

where y_t is a k-vector, μ_0 and μ_1 are intercept vectors of dimension k, $\mathbf{\Phi_j}$ and $\mathbf{\Psi_j}$ are $k \times k$ parameter matrices and

$$
\mathbf{G}(\mathbf{z_t}, \gamma, \mathbf{c}) = diag\{G(z_{1t}, \gamma_{,1}, c_1), \ldots, G(z_{kt}, \gamma_{,k}, c_k)\}
$$

is the diagonal matrix of transition functions. Details of this model can be found in Granger et al. (2010). Applications of the LVSTAR model are generally carried out in the bivariate case (see Camacho (2004)). Within the multivariate framework an error correction model with short-run smooth transition dynamics can also be considered; an example of application in business cycle analysis can be found in Chinn et al. (2014).

13.5.4 Applications

The first work related to business cycle analysis using STAR modelling is the one by Teräsvirta and Anderson (1992). Their model allows the business cycle indicator to alternate between two distinct regimes which represent the two phases of the business cycle, assuming that the transition between these regimes is smooth. They use quarterly industrial production indices from 1960 to 1986 for a set of 13 OECD countries. Lundbergh et al. (2003) examine the data set compiled by J. Stock and M. Watson consisting in 214 monthly US macroeconomic time series, from 1959 to 1996. They study existence of non-linearity and structural changes through TV-STAR framework in order to improve the previous work of Stock and Watson (1999) based on STAR models. Franses and van Dijk (2005) consider quarterly seasonally unadjusted industrial production volume indexes for 18 OECD countries from 1960 to 2002. They forecast those series using linear and different STAR-type models including or not deterministic seasonal components. Their forecasting performance varies across series and horizons. Non-linear STAR models using elaborate seasonal components performs in means for long term horizon. Fok et al. (2005) investigate quarterly growth rates in seasonally adjusted industrial production for 19 main manufacturing sectors over the period 1972 - 2003. They estimate 19 STAR models for each of the 19 sectors, then they consider a kind of aggregate panel STAR models in order to make the sectors comparable. Last, Camacho (2004) uses a bivariate LVSTAR model to assess the non-linear relationship between US GDP and the Conference Board leading indicator.

13.6 Evaluation of models for construction of CCIs

From a practical point of view, it seems legitimate to ask what would be the optimal choice among all those previous models presented in this chapter when the objective is to construct a composite cyclical indicator to monitor economic cycles of a given economy. Taking a pure statistical view, to our knowledge, there is no powerful test able to discriminate between various non-linear alternatives. Some authors have compared non-linear models from a forecasting point of view, we refer for example to the works of Clements and Krolzig (1998), Stock and Watson (1999), Marcellino (2005) or Ferrara et al. (2015). However, comparisons of non-linear models based on the ability to reproduce economic cycles are quite rare. This kind of comparison would be typically based on the Quadratic Probabilistic Score (QPS, see Diebold and Rudebusch (1989)) criterion defined by:

$$QPS(\beta) = \frac{1}{T} \sum_{t=1}^{T} (\hat{P}_t - r_t)^2, \tag{13.1}$$

where $\hat{P}_t$ is the estimated probability of being in the low phase of the cycle, stemming from one given model, and r_t is the reference chronology of cyclical turning points that takes the value one during a low phase of the cycle and zero otherwise. This criterion enables to assess the concordance degree between the indicator and a given reference dating chronology.

Alternatively, a comparison of models can be carried out based on a measure of false and missed signals. By reference to a benchmark chronology, one can compute the number of times that a given phase of the cycle is detected or not by the model. If a given phase is not detected by the model thus it has to be counted as a missed signal. This allows to assess the so-called *type-two error*. Symmetrically, a model can send a signal of turning point while there is no turning point in the benchmark chronology. This kind of event is defined a false signal. By counting those events, one can assess the so-called *type-one error*. Obviously, a performing model will tend to minimize type-one and type-two errors.

Comparing model performances based on type-one and type-two errors strongly relies on the way a specific model is supposed to send a signal of a turning point. Typically, a signal is emitted as soon as $\hat{P}_t$, the estimated probability of being in a specific phase of the cycle, crosses a critical threshold. In general, the natural critical value is 0.5. However, in order to get a more reliable signal, one can think to use a higher value, such as 0.6 or 0.8 for example. In this case we secure the signal by minimizing the risk of false signal. At the same time, the use of a more conservative threshold comes at the cost of sending a more delayed threshold and increasing the risk of missed signal. Here we highlight the trade-off between timeliness and accuracy that practitioners have to keep in mind. From a practical point of view, one has to decide the risk on which to put more emphasis.

To circumvent this issue, alternatives can be envisaged. For example, it may be useful to determine a non-decision region instead of a single threshold. Indeed, it could be decided that between 0.4 and 0.6, no signal is emitted. Obviously, the critical bounds of this region have to be set based on past experience, for example by carrying out a real-time analysis. Also, a censoring rule can be implemented in order to minimize the risk of false signal. Typically, it can be said that the probability has to cross the critical threshold some consecutive times before sending a signal. This may be useful to smooth a signal in the presence of strong volatility.

An example of such considerations can be found in the work by Billio et al. (2013) in which the authors compare the real-time ability of TAR and MS models to detect the euro area business cycles based on vintages of data for both IPI and unemployment rate. These two series are of major importance for short-term economic outlook, as they are usually considered to date economic cycles and are available for the euro area on a monthly basis over a long period of time. The main difference between these two types of models relates to the transition variable used to describe switches: an observed exogenous or endogenous variable for TAR modelling but a latent Markov chain for MS modelling. After a specification step to determine the best modelling in each class, the authors increase successively the original dataset by moving through all the available vintage monthly releases. They re-estimate the models at each step in order to detect in real-time

occurrence of turning points which can characterize changes in the economic phases.

It turns out that the threshold model is very reactive but very volatile and provides too many false signals of recession. Even when using a censoring rule to reduce the number of false signals, the threshold model tends to send more false signals of recessions than the MS model. Each model provides a part of the analysis which can be strengthened by the other one as soon as the key instrument to detect a change in the data is not the same: one explains changes endogenously while the other one exogenously. Another interesting model comparison of non-linear models has been carried out by Deschamps (2008) who compares STAR and MS models to describe the US unemployment rate. Though this latter paper mainly focuses on forecasting, the author points out that the estimated probabilities of being in recession are quite similar, although the MS model appears better able to anticipate turning points.[2]

When dealing with binary response models, it turns out that the role of the reference dating is crucial. As regards the US economy, there is no strong criticism about the NBER dating of recessions. But if researchers want to work on US growth or acceleration cycles, the dating chronology is less clear-cut. As regards the euro area as a whole, uncertainty on recession dates is higher. For example, if we compare CEPR (2003) and CEPR (2009) with Eurostat (Mazzi and Savio (2007)) turning point chronologies, we note that the main difference concerns the second oil shock insofar as CEPR identifies a single recession period from 1980 to 1982 while Eurostat is rather in favour of a double-dip scenario with two recession phases in 1980 and 1981-82. Therefore, the first choice will tend to increase the average duration of a recession phase implying thus different parameter estimates for the model. This example points out one of the limit of the use of binary response models that need turning point dates with a strong confidence level. There is also another practical limitation to the use of binary response models in real-time insofar as their estimation requires exact knowledge of the regime state of the economy for every observation in the estimation period so as to assign values to the dependent variable (r_t). Thus in real-time estimation, when the practitioner wants to update every month the estimation of the model as new economic indicators are available, this limitation becomes quite problematic.

The most comprehensive study that compares binary response models with non-linear alternatives based on Markov-Switching modelling, in terms of dating and forecasting the turning points of the US business cycle, has been carried out by Layton and Katsuura (2001). They use monthly data from 1949 to 1999 stemming from economic indicators computed by the ECRI. The various estimated models are compared according to the classical QPS criterion that measures the adequation between the dating chronology stemming from the estimated models and the one provided by the NBER. On the basis of this within-sample measure, as well as out-of-sample turning point forecasting, the authors conclude that the Markov-Switching specification performs relatively better than other models.

Using those non-linear models presented in this chapter in order to anticipate turning points with a certain lead is quite challenging. Empirical research has put forward the leading role of term spread (i.e. the difference between long-term and short-term interest rates) to predict recessions. However, empirical results are rather mitigated and it is not clear if failures in recession forecasting are due to the data or to the model. The stylised fact that clearly appears is that, in spite of a clear forecasting power, the lead of the term spread is not stable overtime. This lack of stability has been pointed out by Stock and Watson (2003) who note that there is no guarantee that a financial indicator predicting well during a given period of time will still be accurate in later periods. In addition, Giacomini and Rossi (2006) find evidence of breakdowns in the forecasting ability of the US yield curve in predicting US output growth, related to the Fed's monetary regimes. However Bellégo and Ferrara (2012) show that by using a large number of financial variables, the relationship between financial sector and economic activity can be stabilized. Filardo (2004) shows evidence of mitigated results when using a probit model to forecast the 2001 US recession by comparison with the simple rule of thumb based on the CLI or with a regime-switching approach (Neftçi (1982)). Filardo also notes a sensitivity of the results to the use of a logit model with real-time data. Real-time assessment of recession forecasts should be more

---

[2]The author recognizes that this comparison is a bit biased in the sense that the MS probability is based on the whole sample while the STAR probability uses only the latest data observations

systematic in empirical papers, specially when dealing with revised data. Last, Hamilton (2011) concluded that, for various reasons, it was not possible to anticipate recessions. The best that a researcher can do is to detect recession just after they occur.

Overall, from this survey of the literature on parametric models, it turns out that it may be preferable to use models in a complementary way for dating and detecting business cycles. However, if one has to promote one single model to develop CCIs, it turns out that MS models can be recommended. Those model are now well known from practitioners, are easy to implement and have proved useful in this framework. Binary response models are also widely known by practitioners but suffer from a great drawback for real-time analysis as parameters cannot be updated. Last, we note that TAR and STAR models are not yet in the toolkit of business cycles analysts but present a great potential for future work.

Bibliography

Anas J., Ferrara L. (2004), Turning points detection: The ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, No. 1, Vol. 2, pp. 1-36.

Anas J., Billio M., Ferrara L. and Mazzi G.L. (2008), A system for dating and detecting turning points in the euro area, The Manchester School, 76, 5, 549-577.

Attah-Mensah J., Tkacz G. (1998), Predicting Canadian recessions using financial variables: A probit approach, Bank of Canada, Working Paper, No. 98-5.

Bec F., Bouabdallah O., Ferrara L. (2015), Comparing the shape of recoveries: France, the UK and the US, Economic Modelling, 44 (C), 327-334.

Bec F., Bouabdallah O. Ferrara L. (2014), The way out of recession: A forecasting analysis for some euro area countries, International Journal of Forecasting, 30, 3, 539-549.

Bellégo C., Ferrara L. (2009), Forecasting euro area recessions using time-varying binary response models for financial variables, Working Paper, Banque de France, No. 259.

Bellégo C., Ferrara L. (2012), Macro-financial linkages and business cycles: A factor-augmented probit approach, Economic Modelling, 29, 1793-1797.

Bernard H., Gerlach S. (1996), Does the Term Structure Predict Recessions? The International Evidence, Working Paper No. 37, Bank for International Settlements.

Bernard H., Gerlach S. (1998), Does the term structure predict recessions? The international evidence, International Journal of Finance and Economics, 3: 195-215.

Billio M.,Caporin M., Cazzavillan G. (2008), Dating euro15 monthly business cycle jointly using GDP and IPI, Journal of Business Cycle Measurement and Analysis, 3, 3, 333-366.

Billio M., Casarin R. (2010), Identifying business cycle turning points with sequential Monte Carlo methods: An on-line and real time application to the euro area, Journal of Forecasting, 1-2, 145-167.

Billio M., Casarin R. (2011), Beta Autoregressive transition Markov-Switching models for business cycle analysis, Studies in Nonlinear Dynamics and Econometrics, 15, 4.

Billio M., Casarin R., Ravazzolo F., vna Dijk H.K. (2012), Combination schemes for turning point predictions, Quarterly Review of Economics and Finance, 52, 402-412.

Billio M., Ferrara L., Guégan D., Mazzi G.L (2013), Evaluation of Nonlinear time-series models for real-time business cycle analysis of the Euro area, Journal of Forecasting, 32, 7, 577-586.

Boldin M. D. (1996), A check on the robustness of Hamilton's Markov Switching model approach to the economic analysis of the business cycle, Studies in Nonlinear Dynamics and Econometrics , vol. 1, pp. 35-46.

CEPR (2003), Press Release, Euro Area Business Cycle Dating Committee, September 22.

CEPR (2009), Press Release, Euro Area Business Cycle Dating Committee, March 24.

Camacho M. (2004), Vector smooth transition regression models for US GDP and the composite index of leading indicators, Journal of Forecasting, 23, 173-196.

Camacho M., Perez-Quiros G., Poncela P. (2014), Green shoots and double dips in the euro area: A real time measure, International Journal of Forecasting, 30, 3, 520-535.

Camacho M., Perez-Quiros G., Poncela P. (2012), Markov-Switching dynamic factor models in real-time, Working Paper 1205, Banco de España.

Carrasco M., Hu L., Ploberger W. (2009), Optimal test for Markov Switching parameters, Working Paper, University of Montreal, Canada.

Chan K.S. (1993), Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, Annals of Statistics, 21, 520-533.

Chen Z., Iqbal A., Lai H. (2011), Forecasting the probability of US recessions: a Probit and dynamic factor modelling approach, Canadian Journal of Economics, 44, 2, 652-671.

Chauvet M. (1998), An economic characterization of business cycle dynamics with factor structure and regime switches, International Economic Review, 39, 969 - 996.

Chauvet M., Hamilton J.D. (2006), Dating business cycle turning points in real time, Nonlinear Time Series Analysis of Business Cycles, D. van Dijk, C. Milas and P. Rothman (eds.), 1-54, Contribution to Economic Analysis, Elsevier.

Chauvet M., Potter S. (2002), Predicting recessions: Evidence of the yield curve in the presence of structural breaks, Economics Letters, 77,2, 245-253.

Chauvet M., Potter S. (2005), Forecasting recessions using the yield curve, Journal of Forecasting, 24, 2, 77-103.

Chauvet M., Potter S. (2010), Monitoring business cycles with structural changes, International Journal of Forecasting, 26, 4, 777-793.

Chauvet M., Piger J. (2008), A comparison of the real-time performance of business cycle dating methods, Journal of Business Economics and Statistics, 26, 1, 42-49.

Chinn M., Ferrara L., Mignon V. (2014), Explaining US employment growth after the Great Recession: The role of output-employment non-linearities, 42, C, 118-129.

Clements M. P., Krolzig H.M. (1998), A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, Econometrics Journal, 1, C47-C75.

Darné O., Ferrara, L. (2011), Identification of slowdowns and accelerations for the Euro area economy, Oxford Bulletin of Economics and Statistics, 73, 3, 335-364.

Deschamps P. (2008), Comparing smooth transition and Markov switching autoregressive models of US unemployment, Journal of Applied Econometrics, 23, pp. 435-462.

Diebold F.X., Lee J.H., Weinbach G. (1994), Regime switching with time-varying transition probabilities, in C. Heargraves, Nonstationary time series and cointegration, Oxford.

Diebold F.X., Rudebusch G.D. (1989), Scoring the leading indicators, The Journal of Business, 62, 3, pp. 369-391.

Diebold F.X., Rudebusch G.D. (1996), Measuring business cycles: A modern perspective, Review of Economics and Statistics, 78, 1, 67 - 77.

van Dijk D., Franses H.P. (1999), Modelling multiple regimes in the business cycle, Macroeconomics Dynamics, 3, 311-40.

van Dijk, D., Teräsvirta T., Franses H.P. (2002), Smooth transition autoregressive models - A survey of recent developments, Econometric Reviews, 21, 1, 1-47.

Duarte A., Venetis I.A., Paya I. (2005), Predicting real growth and the probability of recession in the Euro-area using the yield spread, International Journal of Forecasting, 21, 2, 261-277.

Dueker M. (1997), Strengthening the case for the yield curve as a predictor of US recessions, Federal Reserve Bank of St. Louis, Review, vol. 79, No. 2, pp. 41-51.

Dueker M. (2002), Regime-dependent recession forecasts and the 2001 recession, Federal Reserve Bank of St. Louis, Review, pp. 29-36, November/December 2002.

Durland J.M., McCurdy T.H. (1994), Duration-dependent transitions in a Markov model of US GNP growth, Journal of Business and Economic Statistics, 12, 279-288.

Eitrheim, O., Terasvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models, Journal of Econometrics, 1996, 74, 1, pages 59-75, available online at =https://ideas.repec.org/a/eee/econom/v74y1996i1p59-75.html

Enders W., Falk B., Siklos P. (2007), A threshold model of real U.S. GDP and the problem of constructing confidence intervals in TAR models, Studies in Nonlinear Dynamics and Econometrics, 11, 3.

Estrella A. (1998), A new measure of fit for equations with dichotomous dependent variables, Journal of Business and Economic Statistics, 16, 2, 198-205.

Estrella A., Hardouvelis G. (1991), The term structure as a predictor of real economic activity, Journal of Finance, vol. 46, pp. 555-578.

Estrella A., Mishkin F.S. (1995), The Term Structure of Interest Rates and Its Role in Monetary Policy for the European Central Bank, National Bureau of Economic Research Working Paper no. 5279, September.

Estrella A., Mishkin F.S. (1998), Predicting US recessions: Financial variables as leading indicators, Review of Economics and Statistics, vol. 80, pp. 45-61.

Estrella A., Rodrigues A.P., Schich S. (2003), How stable is the predictive power of the yield curve? Evidence from Germany and the United States, Review of Economics and Statistics, 85, 3, 629-644.

Ferrara L. (2003), A three-regime real-time indicator for the US economy, Economic Letters, 81, (3), 373 - 378.

Ferrara L. (2009), Caractérisation et datation des cycles économiques en zone Euro, Revue Economique, Vol. 60, No. 3, pp. 703-712.

L. Ferrara and D. Guegan (2005) Detection of the industrial business cycle using SETAR models, Journal of Business Cycle Measurement and Analysis , Vol. 2, No. 3, pp. 353-372.

Ferrara L., Guégan D. (2006), Detection of industrial business cycles using SETAR models, Journal of Business Cycle Measurement and Analysis, 2, (3), 1-20.

Ferrara L., Marcellino M., Mogliani M. (2012), Macroeconomic forecasting during the Great Recession: The return of non-linearity?, International Journal of Forecasting, 2015.

Filardo A.J. (1994), Business cycle phases and their transitional dynamics, Journal of Business and Economic Statistics, 12, 299-308.

Filardo A.J. (1999), How reliable are recession prediction models?, Federal Reserve Bank of Kansas City, Quarterly Review , pp. 35-55.

Filardo A.J. (2004), The 2001 US recession: what did recession prediction models tell us?, BIS Working Papers No 148.

Filardo A.J., Gordon S.F. (1994), Business cycle durations, Journal of Econometrics, vol. 85, n° 1, pp. 99-123.

Filardo, A.J., Gordon S.F. (1998), Business cycles duration, Journal of Econometrics, 85, 1, 99-123.

Fok D., van Dijk D., Franses P.H.(2005), Forecasting Aggregates using Panels of Nonlinear Time Series, International Journal of Forecasting, 21(4), pp 785-794.

Forni M., Hallin M., Lippi M., Reichlin L. (2005), The generalized dynamic factor model: One-sided estimation and forecasting, Journal of the American Statistical Association, 100, 830-840.

Franses P.H., van Dijk D. (2000), Non-Linear Time Series Models in Empirical Finance, Cambridge University Press, Cambridge.

Franses P.H., van Dijk D. (2005), The forecasting performance of various models for seasonality and nonlinearity for quarterly industrial production, International Journal of Forecasting, Elsevier, 21, 1, pages 87-102.

Garcia R. (1998), Asymptotic null distribution of the Likelihood Ratio test in Markov-Switching models, International Economic Review, 39, 763-788.

Giacomini R., Rossi B (2006), How stable is the forecasting performance of the yield curve for output growth?, Oxford Bulletin of Economics and Statistics, 68, 783-795.

Gonzalo J., Pitarakis J.Y. (2002), Estimation and model selection based inference in single and multiple threshold models, Journal of Econometrics, 110, 319 - 352.

Granger C.W., Teräsvirta T., Tjostheim D. (2010), Modelling Nonlinear Economic Time Series, Oxford University Press, Oxford.

Hamilton J.D. (1988), Rational expectations econometric analysis of change in regime: An investigation of the term structure of interest rates, Journal of Economic Dynamics and Control, 12, 385-423.

Hamilton J.D. (1989), A new approach to the economic analysis of non-stationary time series and the business cycle, Econometrica, 57, 357-384.

Hamilton J.D. (1990), Analysis of time series subject to changes in regime, Journal of Econometrics, 45, 39 - 70.

Hamilton J.D. (1994), Time Series Analysis, Princeton University Press.

Hamilton J.D. (1996). Specification Testing in Markov-Switching Time-Series Models, Journal of Econometrics, Jan. 1996.

Hamilton J.D. (2003), What is an oil shock?, Journal of Econometrics, 113, 363-398.

Hamilton J.D. (2011), Calling Recessions in Real Time, International Journal of Forecasting October-December 2011, vol. 27, no. 4, pp. 1006-102.

Hansen B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, Econometrica, 64, 2, 413-430.

Hansen B.E. (1997), Inference in TAR Models, Studies in Nonlinear Dynamics and Econometrics, 2, 1-14.

Hansen B.E. (1999), Testing fo linearity, Journal of Economic Surveys, 13, 551-576.

Hansen B.E., Seo B. (2002), Testing for two-regime threshold cointegration in vector-error correction models, Journal of Econometrics, 110, 293-318.

Kapetanios G. (2000), Small sample properties of the conditional least squares estimator in SETAR models, Economic Letters, 69, 3, 267-276.

Kauppi H., Saikkonen P. (2008), Predicting US recessions with dynamic binary response models, The Review of Economics and Statistics, 90, 4, 777-791.

Kim C.J., Nelson C.R. (1998), Business cycle turning points, a new coincident index, and tests of duration dependence based on dynamic factor model with regime-switching, The Review of Economics and Statistics, 80, 188-201.

Kim C.J., Nelson C.R. (1999), State-Space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications, Cambridge, MA: MIT Press.

Kim C.J., Morley J., Piger J. (2005), Nonlinearity and the Permanent Effects of Recessions, Journal of Applied Econometrics 20.

King T.B., Levin A.T., Perli R. (2007), Financial market perceptions of recession risk, FEDS Working Paper No. 2007-57.

Krolzig H.M. (1997), Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis, Lecture Notes in Economics and Mathematical Systems, New York/Berlin/Heidelberg: Springer.

Krolzig H.M. (2003), Constructing turning point chronologies with Markov-switching vector autorgressive models: the Euro-zone business cycle, in Eurostat Colloquium on Modern Tools for Business Cycle Analysis.

Layton A., Katsuura M. (2001), Comparison of regime switching, probit and logit models in dating and forecasting US business cycles, International Journal of Forecasting, 17, 403-417.

Layton A.P., Smith D.R. (2007), Business cycle dynamics with duration dependence and leading indicators, Journal of Macroeconomics, 29, 855-875.

Lundbergh S., Teräsvirta T., van Dijk D. (2003), Time-varying smooth transition autoregerssive models, Journal of Business and Economic Statistics, 21, 104-121.

Luukkonen R., P. Saikkonen and T. Teräsvirta (1988), Testing linearity against Smooth-Transition Autoregressive models, Biometrika, 75, 3, 491-499.

Marcellino M. (2005), Instability and non-linearity in the EMU, in Nonlinear Time Series Analysis of Business Cycles, C. Milas, P. Rothman and D. van Dijk (eds.), Elsevier, Amsterdam.

Mazzi G.L., G. Savio (2007), Growth and Cycle in the Euro-Zone, Palgrave MacMillan, 02/2007

McConnell M., Perez-Quiros G. (2000), Output fluctuations in the United States: What has changed since the early 1980s, American Economic Review, 90,5, 1464-1476.

Moneta F. (2005), Does the yield spread predict recessions in the Euro area?, International Finance, 8, 2, 263-301.

Montgomery A., Zarnowitz V., Tsay R., Tiao G. (1998), Forecasting the U.S. unemployment rate, Journal of the American Statistical Association, 93, 442, pp. 478-493.

Nalewaik J.J. (2012), Estimating Probabilities of Recession in Real Time Using GDP and GDI, Journal of International Money and Finance, 44, 1, 235-253.

Neftçi S. (1982), Optimal predictions of cyclical downturns, Journal of Economic Dynamics and Control, 4, 307-327.

Neftçi S. (1984), Are economic time series asymmetric over the business cycle, Journal of Political Economy, 92, 307-328.

Nyberg H. (2010), Dynamic probit models and financial variables in recession forecasting, Journal of Forecasting, 29, 1-2, 215-230.

Potter S.M. (1995), A nonlinear approach to US GNP, Journal of Applied Econometrics, 10, 109-125.

Proietti T. (1998), Characterizing asymmetries in business cycles using smooth-transition structural time series models, Studies in Nonlinear Dynamics and Econometrics, 3, 141-156.

Quandt R.E. (1958), The Estimation of Parameters of Linear Regression System Obeying Two Separate Regime, Journal of the American Statistical Association, 55, 873-880.

Rudebusch G., Williams J. (2009), Forecasting recessions: the puzzle of the enduring power of the yield curve, Journal of Business and Economics Statistics, 27 (4), 492-503.

Sensier M., van Dijk D. (2004), Testing for changes in volatility of US macroeconomic time series, Review of Economics and Statistics, 86, 833-839.

Sichel D. (1991), Business Cycle Duration Dependence: A Parametric Approach, The Review of Economics and Statistics, 73, 254-260.

Sichel D. (1994), Inventories and the Three Phases of the Business Cycle, Journal of Business and Economic Statistics, 12, 269-277

Stock J., Watson M. (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R. F. and White, H. (eds.), Cointegration, causality and forecasting. A festschrift in honour of Clive W. J. Granger. Oxford University Press: Oxford (UK).

Stock J., Watson M. (2002), Macroeconomic forecasting using diffusion indexes, Journal of Business and Economic Statistics, 20, 147-162.

Stock J., Watson M. (2003), Forecasting output and inflation: The role of asset prices, Journal of Economic Literature, 41, 3, 788-829.

Strikholm B., Teräsvirta T. (2006), A sequential procedure for determining the number of regimes in a threshold autoregressive model, Econometrics Journal, 9, 472-491.

Teräsvirta T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, Journal of the American Statistical Association, 89, 208 - 218.

Teräsvirta T., Anderson H. (1992), Characterizing nonlinearities in business cycles using Smooth Transition Autoregressive Models, Journal of Applied Econometrics, 7, S. 119-S. 136.

Tiao G.C., Tsay R.S. (1994), Some advances in non-linear and adaptive modelling in time series, Journal of Forecasting, 13, 109 - 131.

Tong H. (1978), On a threshold model, in Pattern Recognition and Signal Processing, (ed.) C.H. Chen, Amsterdam: Sijhoff and Noordhoff.

Tong H. (1990), Non-linear Time Series: A Dynamical Approach, Oxford Scientific Publications, Oxford.

Tong H., Lim K.S. (1980), Threshold autoregression, limit cycles and cyclical data, Journal of the Royal Statistical Society, B, 42, 245-292.

Tsay R.S. (1989), Testing and modeling threshold autoregressive processes, Journal of the American Statistical Association, 84, 231-240.

Tsay R.S. (1998), Testing and modeling multivariate threshold models, Journal of the American Statistical Association, 93, 1188-1202.

Wooldridge J. (1994), Estimation and Inference for Dependent Processes, Cambridge University Press, Cambridge.

14

Cyclical Composite Indicators Detecting Turning Points

Authors

Jacques Anas
Independent Expert
jacques.anas@free.fr

Laurent Ferrara
Banque de France
laurent.ferrara@banque-france.fr

Leonardo Carati
GRETA
leonardo.carati@gmail.com

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Monica Billio
Ca' Foscari University of Venice
billio@unive.it

Table of Contents

14.1 Introduction

The problem of detecting turning points in the business cycles in a timely manner is an important issue with the grave impact for the policy makers. Quick and reliable recognition of possible slowdown or recession allows for a quick response in terms of fiscal and monetary policy. However, there is always a trade-off between speed and accuracy. Therefore, in constructing the indicators, we focused on reliability rather than flagging every decrease in the economic growth. Out of the three types of indicators used in the turning point detection: leading – used to predict, coincident – used to detect and lagging – used to confirm, we have chosen the coincident ones. This chapter describes three coincident indicators, proposed in a few versions, which can be used in detection of the turning points. They were introduced already in the literature, but we believe that a more exhausting description could be beneficial. Also, the indicators were tested "in field" in the last years and developed further in order to incorporate recent findings. In order to make the indicators and turning points easier to be understood we first introduce the concept of three types of fluctuations, which are usually discussed in the literature: classical business cycle, growth cycle and acceleration (also referred to as the growth rate) cycle, showing their interweaving nature in the $\alpha AB\beta CD$ approach, which was initially proposed by Anas and Ferrara (2004a).

Each characterization of cyclical movements has obviously advantages and drawbacks, and relates to specific aspects of economic cycles. We also show the chronology of the past business cycles to illustrate how the cycles looked like and in order to give real-life examples. As practice shows, real series quite often follow a very intriguing pattern stemming out from the economic conditions creating the cycles of various lengths. They are described in detail in the Section 14.2. The dating exercise is however not as easy as it seems to be, as it has to be based on sufficiently long time-series covering several cycles. Unfortunately this requirement is difficult to fulfill because statistics can be affected by several methodological changes, evolving statistical aggregates and classifications etc., which will inevitably shorten their length (or cause breaks). Finally, in the process of dating and detecting turning points, statistical findings need to be interpreted and validated from an economic and even more, from a political point of view. Therefore, the analysis of turning point usually requires a complex and thorough approach.

The next parts of the chapter are arranged in the following way: Section 14.2 explains and illustrates the $\alpha AB\beta CD$ approach, giving theoretical background to the subsequent analysis; Section 14.3 shows methodology for dating chronology and gives examples for the euro area; Section 14.4 as the most bulky part, introduces the indicators, first in the univariate approach and then in the multivariate one. The indicators are described in details, including methodology, component variables and models used to develop them. At the end a set of indicators for the euro area is shown as an example. Section 14.5 concludes, an annex is provided in section 14.6.

14.2 The extended $\alpha AB\beta CD$ approach of economic cycles

In this section we describe the empirical approach developed by Eurostat to monitor economic cycles turning points, referred to as the extended $\alpha AB\beta CD$ approach. This approach is based on the classification of economic cycles into business, growth and acceleration cycles. Each type of cycle possesses its own turning points and there exists a sequence of turning points. Here, we recall the definition of those cycles and define the various turning points.

14.2.1 The $ABCD$ approach for business and growth cycles

In the literature on business cycle analysis, the studies generally refer to the business or growth cycles. Basically, the business cycle refers the (log-)level of the series, as defined by Burns and Mitchell (1946). Turning points of the business cycle delimit periods of recessions (negative growth rate) and expansions (positive growth rate). The business cycle is characterized by strong asymmetries in its phases, concerning for example durations or amplitudes. For example, since 1970, the average duration of an expansion phase in the euro area varies between 8 and 11 years according to the studies while the average duration of a recession is only of one year. It seems also that only recessions possess the property of duration-dependence implying thus that the probability of switching to the regime of expansion increases with time. The growth cycle, introduced by Mintz (1969), is the cycle of the deviation to the long-term trend, which can be seen as the potential or tendencial growth. This cycle is sometimes referred to as the output gap. Both business and growth cycles have been widely studied in the literature (Artis et al. (2004), Artis et al. (2004), Anas and Ferrara (2004a), Zarnowitz and Ozyildirim (2006), Anas et al. (2007a), Anas et al. (2007b)).

We recall here briefly the ABCD approach proposed by Anas and Ferrara (2004a) and in use at Eurostat. Specific turning points are associated with business and growth cycles. Points B and C will be the extreme points of the classical cycle, while points A and D will be those of the growth cycle (see Figure 14.1). The ABCD approach is based on the four following principles:

- The turning point dating or detection issue must be considered as a progressive follow-up of the cyclical movement. Even if no cycle is similar to the previous one, the sequence of turning points is always respected in practice. A slowdown movement will first materialize in a peak in the growth cycle (point A) and, if it is getting worse, the growth rate will become negative (point B) implying thus a recession. For an upward movement, the sequence will be a trough in the business cycle (point C) and a recovery of the growth rate above the trend growth rate (point D).

- If the slowdown does not gain in intensity to become a recession, then point A will not be followed by point B. In other words, the economy can experience a descending phase of the growth cycle (peak A and trough D) without going through a recession (peak B and trough C). This is for example what happened between 1999 and 2003 for the euro area. the temporal sequencing of those points (A and B for peaks and C and D for troughs) the $ABCD$ strategy for turning points analysis.

- It is worth noticing that the $ABCD$ approach is an empirical one. The empirical analysis we propose does not rely on any theoretical approach of the nature and the causes of the cycles. Therefore, it cannot be seen as a proposal for an unified theory that applies to both business and growth cycles alike. This is rather a data-driven approach that enables to provide successive real-time signals to decision makers in terms of turning points. There are different patterns for cyclical evolutions. A recession may occur suddenly so that A and B would coincide. Symmetrically, in a rapid exit of a recession, C and D would coincide. As regards the CD phase, the economy can go from C to D either with a fast pace (V-shaped exit, the dates of C and D are thus close) or with a slow pace (e.g. *jobless recovery*, the dates of C and D are distant), but D will always be the date where the deviation to trend reaches a minimum.

- For both dating and real-time detection exercises, business and growth cycles are treated separately, although the ABCD chronology has to be respected.

14.2.2 Extension of the $ABCD$ approach to the acceleration cycle

Beyond the business and growth cycles discussed above, the empirical literature focuses also on the acceleration cycle, sometimes referred to as the growth rate cycle. The acceleration cycle is the cycle described by increases and decreases in the growth rate of economic activity. A turning point of this cycle occurs when a local extremum is reached. This cycle is thus a sequence of decelerating and accelerating phases. Such

Figure 14.1: $\alpha AB\beta CD$ approach: A theoretical example (source: Anas and Ferrara (2004a)). Turning points of the business cycle (middle) are B and C, those of the growth cycle (bottom) are A and D and those of the acceleration cycle (top) are α and β.
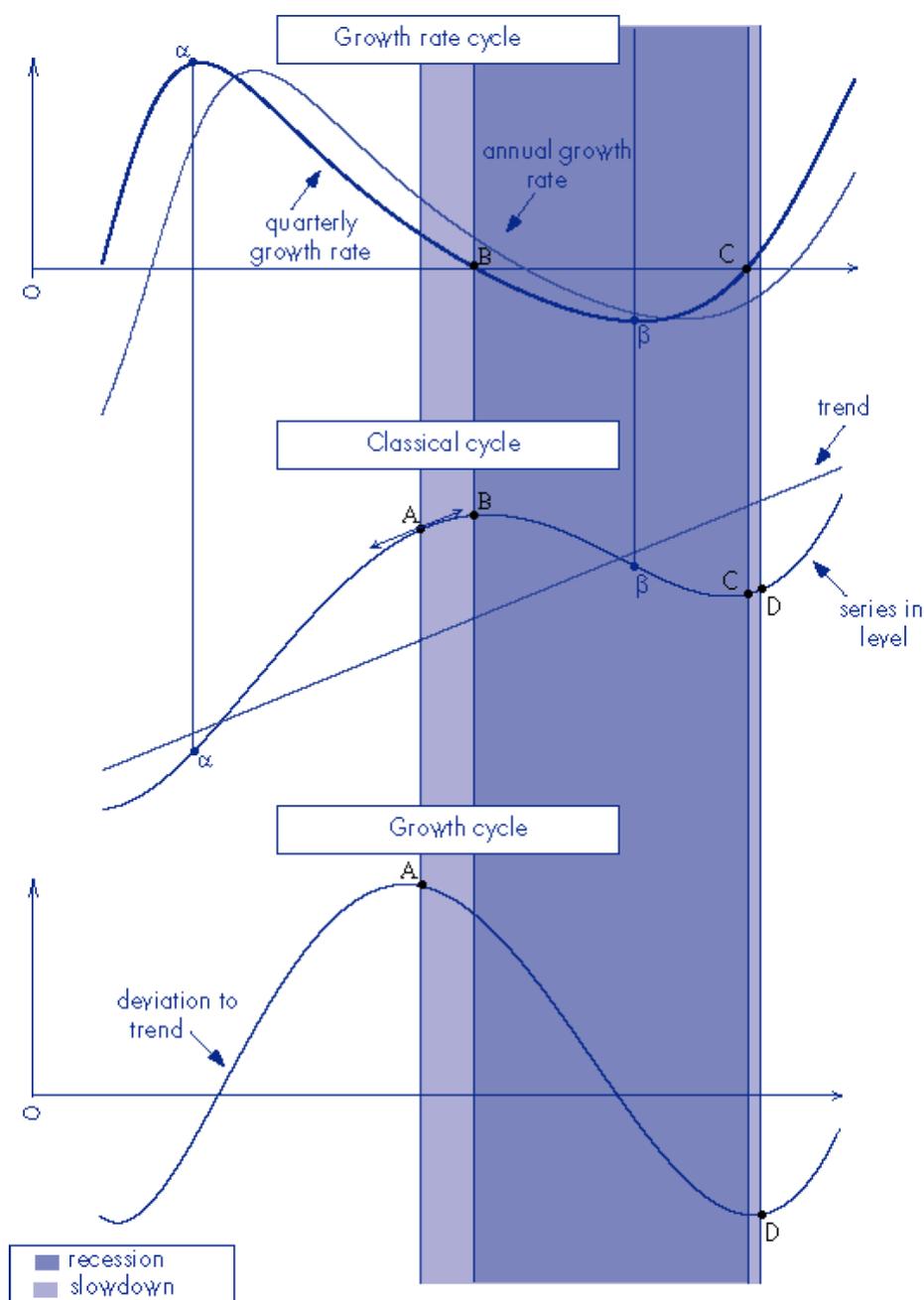
a cycle is very interesting for the short-term economic analysis of the euro area, not often affected by recessions, because of its high frequency. Indeed, this high frequency enables to provide with a cyclical diagnosis on a frequent basis. However, its more pronounced volatility implies a more complex real-time detection and it is often uneasy to have a clear economic interpretation for the phases of the acceleration cycle.

The acceleration cycle can be easily integrated into the framework of the ABCD approach. Let denote respectively α and β the peaks and troughs of the acceleration cycle that can be seen on Figure 14.1 (top graph). It is clear that a peak A in the growth cycle is necessarily preceded by a peak in the acceleration cycle α. Obviously, the activity decelerates before its growth rate falls below its tendencial growth rate. However, the

reverse is not true, that is a peak α does not necessarily imply a peak A: the growth rate can increase again without having reached its tendencial value. Following the same principle, concerning the exit of the cycle, a trough β occurs before a trough D, but here again the occurrence of β does not imply necessarily D. This sequence of turning points $\alpha AB\beta CD$, that we define in this chapter as the *extended ABCD approach*, constitutes an useful tool to assess the conjunctural economic fluctuations, both for dating and detecting exercises. For example, during a recession phase, the first optimistic signal will be given at the trough β, where the growth rate of the activity will begin an ascending phase. The exit of the recession will occur lately at point C where the growth rate will become positive.

14.2.3 Examples

In order to illustrate how the cycles can look, the following graph shows the theoretical cases, which could be encountered in empirical analysis. The examples are of course exaggerated and are not likely to exist in the real economy. There are two graphs - the first one, Figure 14.2, showing the growth rate and the second one, Figure 14.3, showing the corresponding GDP value.

Figure 14.2: Growth rate: Illustrative example
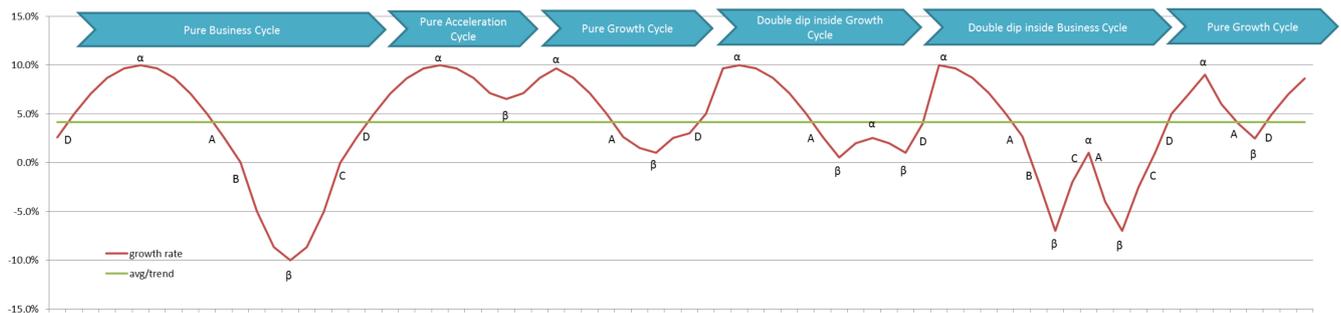
Figure 14.3: GDP: Illustrative example
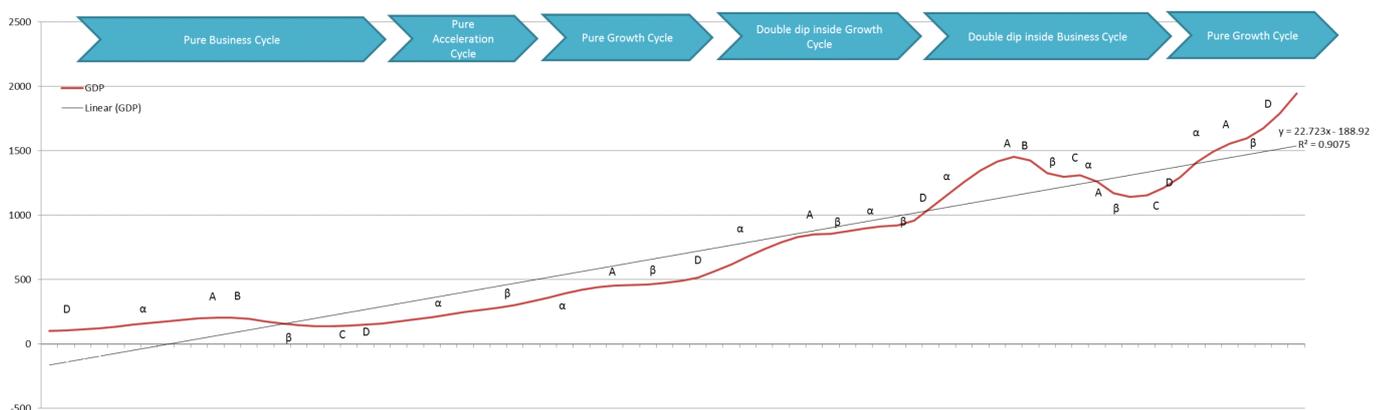
The timelines of above graphs are synchronized to show how the turning points and the $\alpha AB\beta CD$ sequences are exactly the same in both. The real sequences are unfortunately less straightforward and turning points are not that obvious to spot.

14.3 Dating chronologies

Based on the previous $\alpha AB\beta CD$ approach, the first step towards the construction of a composite cyclical indicator for turning points is the construction of a reference chronology of cyclical turning points. In this section, we point out the need for such dating chronology, we briefly describe the tools available for its construction then we provide the euro area chronologies as an example.

14.3.1 Why is it important?

The need for a cyclical turning point chronology is now widely recognized by experts and practitioners of economic analysis. As an example of application, it may help to compare the cycles between nations or to point out links between the cycles of various economic aggregates. However, it turns out that the most important use of the turning point chronology consists in establishing a reference cycle dating for a given country or economic area. Indeed, this reference cycle is often used in empirical studies either to classify economic series according to their properties (leading, coincident or lagging) or to validate real-time detection and forecasting methods. It is obvious that dating is an ex post exercise. In this respect, accuracy is a more important criterion than timeliness. Because of the lack of timeliness, dating may not be useful for economic decision-making. As a matter of fact, governments and central banks are very sensitive to indicators showing signs of deterioration in growth to allow them to adjust their policies sufficiently in advance, avoiding further deterioration or a recession. In this respect, timing is important and the earlier the signal, the better. This issue is linked to the *real-time detection* concept. However, to validate their methods of real-time detection, researchers need a reference turning point chronology for the cycle they aim at tracking. Only the US have a well known benchmark turning point chronology of the business cycle established by the Dating Committee of the NBER.

As regards the euro area, European institutes, such as the CEPR (CEPR (2003), CEPR (2009)) or Eurostat (Mazzi and Savio (2007), or Anas et al. (2008)), have proposed a reference dating for the business cycle. Eurostat proposes also a turning point chronology for the growth cycle integrated with the business cycle. Moreover, the OECD updates regularly a monthly chronology for the growth cycle of the euro area, as well as for its members, available on the institution web site.[1] Otherwise, several academic studies have also developed dating chronologies for both the business and growth cycles, see for example Artis et al. (2004), Artis et al. (2004), Anas and Ferrara (2004a), Mönch and Uhlig (2005), Anas et al. (2007a). A review of the various turning point chronologies can be found in the paper of Anas et al. (2008). A historical turning point chronology of the euro area acceleration cycle has been proposed in Harding (2004), but his analysis ends in 1998, and more recently in Darné and Ferrara (2011).

14.3.2 Methodology

According to the results of Anas and Ferrara (2004b) and Anas et al. (2007a) in the framework of turning points dating chronology, we are in favor of non-parametric procedures instead of parametric ones used for example in Schirwitz (2009) for the German case. Indeed, as regards the dating exercise, which is an *ex post* exercise, it turns out that parametric models tend to be less robust to the sample size. That is when new data arrives, there is a non-null probability to observe changes in estimated parameters, leading thus to changes in the estimated probability of being in a given phase of the cycle. Obviously, revising the chronology through time is something that has to be avoided. Non-parametric methods do not present such drawbacks. In addition, non-parametric approaches can be easily adapted to different series and countries. But it is noteworthy that parametric models (presented in Ferrara and Mazzi (2016), this volume) are more adapted to the real-time detection of turning points in the economic cycles. Generally, a basic version of the non-parametric dating

---

[1]www.oecd.org

algorithm proposed by Bry and Boschan (1971) (BB hereafter) and modified by Harding and Pagan (2002) is implemented. This approach is very simple to handle and has been used in several empirical papers dealing with business cycles analysis (see, for example, Harding (2004), Engel *et al.*, 2005, Anas et al. (2007a) or Demers and MacDonald (2007), Anas et al. (2008), or Darné and Ferrara (2011)). This methodology is often applied to the broadest measure of economic activity, that is GDP, but can be implemented on monthly proxy variables such as industrial production or any synthetic indicator reflecting aggregate activity. From this aggregated dating, some stylized facts of the cycle (duration, amplitude, excess etc.) are also measured to validate this turning point chronology.

Assume $(Y_t)_t$ is the series of interest (GDP or IPI), seasonally adjusted, corrected from trading days and outliers. The basic BB algorithm detects a peak at date t if the following condition is verified:

$$\{(\Delta_k Y_t, \ldots, \Delta_1 Y_t) > 0, (\Delta_1 Y_{t+1}, \ldots, \Delta_k Y_{t+k}) < 0\} \tag{14.1}$$

and detects a trough at date t if the following condition is verified:

$$\{(\Delta_k Y_t, \ldots, \Delta_1 Y_t) < 0, (\Delta_1 Y_{t+1}, \ldots, \Delta_k Y_{t+k}) > 0\}, \tag{14.2}$$

where the operator Δ_k is defined such as $\Delta_k Y_t = Y_t - Y_{t-k}$. Harding and Pagan (2002) suggest $k = 2$ for quarterly data and $k = 5$ for monthly data. Generally, turning points within six months of the beginning or end of the series are disregarded. Last, a procedure for ensuring that peaks and troughs alternate is developed, for example by imposing that in the presence of a double through, the lowest value is chosen and that in the presence of a double peak, the highest value is chosen. Censoring rules related to the minimum duration of phases are also imposed in the original algorithm specifying that a phase must last at least six months and that a complete cycle (from peak to peak) must last at least 15 months. In fact, this censoring rule applies for the business cycle because, as noted by the NBER in its seminal definition, a recession must last *more than a few months*, but there is no reference minimum duration.

Three main characteristics are often invoked in order to identify the phases of a cycle, namely the 3D's (duration, depth and diffusion) or, as in Banerji (1999), the 3P's (persistent, pronounced and pervasive). Persistence (or duration) means that the phase must last more than a few months. Generally, starting from the Bry and Boschan (1971) rule, empirical studies consider that a phase of the cycle must last at least five months. A pronounced phase of a cycle is a phase with a sufficient amplitude (depth) from the peak to the trough and conversely. Last, to be recognized as a phase of the cycle, the cycle must be diffused either across the sectors or across the various countries of an economic area.

Assume that the previous step has produced the same number J of accelerating and decelerating phases. For $j = 1, \ldots, J$, we note D_j^a and D_j^d the durations in months of the jth accelerating and decelerating phases, respectively. The amplitude of a descending (or ascending) phase is measured by the absolute distance between the peak and the trough (or the trough and the peak). We note $A_j = |Y_{t_P} - Y_{t_T}|$ the amplitude of a given phase j, where Y_{t_P} is the growth rate of the series at date of peak and Y_{t_T} is the growth rate at date of trough. To sum up duration and amplitude of a phase j, an index of severity, noted S_j, is often used. The severity is sometimes referred to as the *triangle approximation to the cumulative movements* (Harding and Pagan (2002), p. 370) and is defined by:

$$S_j = 0.5 \times D_j \times A_j. \tag{14.3}$$

The severity index measures the area of the triangle with length D_j and height A_j. In fact, the actual measure of cumulative movements, which may be substantially different from S_j in case of departure from linearity, is given by:

$$C_j = \left| \sum_{i}^{D_j} (Y_i - Y_0) \right| - 0.5 \times A_j, \tag{14.4}$$

where Y_0 is the value of the variable at the date of peak, Y_{t_P}, for a decelerating phase (or at the date of trough, Y_{t_T}, for an accelerating phase). The term $0.5 \times A_j$ removes the bias due to the approximation of a triangle by a sum of rectangles. Consequently, for a given phase j, the difference between the observed growth and a linear growth can be measured by the *excess cumulated movements* index defined by:

$$E_j = (C_j - S_j)/D_j. \tag{14.5}$$

This excess index E_j proposed by Harding and Pagan (2002) can be seen as a measure of the departure to the linearity for the growth rate of a given phase. The excess index is divided by the duration so that phases can be compared, independently from their duration. A null excess index implies a linear growth within a phase (decreasing or increasing growth), thus a constant acceleration (negative or positive). For a descending phase, a positive excess index means that the loss of growth is greater than it would be with a linear growth and a negative index indicates that the loss is lower. For an increasing phase, a positive excess index means that the gain of growth is greater than it would be with a linear growth and a negative index indicates that the gain is lower. We can also refer to Camacho et al. (2008) for a description of the measures of duration, depth and excess and for a bootstrap approach to evaluate the uncertainty associated to these measures.

Assume now that we get at disposal n dating chronologies. Typically, this can be the case for an indirect dating of a monetary union (e.g. euro area) or a trade area (e.g. Asean, European Union etc.). Alternatively, for a single country we can have a decomposition into n sectors. The issue is how to assess the synchronization or the diffusion of cycles among the n components.

A first tool is the following diffusion index D_t defined as follows:

$$D_t = \frac{1}{\sum_{i=1}^{n} \omega_i} \sum_{i=1}^{n} \omega_i R_{it}, \tag{14.6}$$

where ω_i is the weight of the ith component and R_{it} is a binary variable equal to 1 when the component i is in the low phase of the cycle and 0 otherwise. As a decision rule, we use the natural threshold of 0.50 to identify a switch in regimes, namely a turning point in the aggregate.

Second, in order to assess synchronization among the country-specific cycles, the concordance index allows an estimate of the fraction of time that cycles are in the same phase (decelerating or accelerating).[2] Let $(S_{it})_t$ denote the binary variable that represents the phase of the cycle (low phase : $S_{it} = 0$, high phase : $S_{it} = 1$) for a given component i. In the bivariate case, for two components i and j, the concordance index CI can be expressed in this way:

$$CI = \frac{1}{T} \sum_{t=1}^{T} I_t, \tag{14.7}$$

where

$$I_t = S_{it}S_{jt} + (1 - S_{it})(1 - S_{jt}). \tag{14.8}$$

---

[2]See Artis et al. (2004), Artis et al. (2004), Harding and Pagan (2002) or Montoya and de Haan (2008) for others measures of synchronization.

At each date t, for all $(S_{it}, S_{jt}) \in \{0, 1\}$, I_t is equal to 1 when $S_{it} = S_{jt}$ and equal to 0 when $S_{it} = (1 - S_{jt})$. This tool is very interesting in empirical studies to assess the synchronization between two cycles. Anyway, we should keep in mind that the concordance index should be misleading because, even if the correlation between S_{it} and S_{jt} is zero, the concordance index CI is equal to 0.5 only if the mean of S_{it} and S_{jt} are both equal to 0.5. It is possible to prove that the expectation of the concordance index depends on the unconditional probabilities of S_{it} and S_{jt} (see Harding and Pagan (2002); Artis et al. (2004)). For example, if the unconditional probability is close to 0.9, as it is the case for the business cycle, it can be proven that, even though the correlation coefficient between the countries is zero, the expectation of CI is close to 0.84. Thus, this index has to be carefully considered in empirical studies.

Harding and Pagan (2006) propose procedures to test the hypothesis that cycles are either un-synchronized or perfectly synchronized, based on the knowledge of the two binary variables $(S_{it})_t$ and $(S_{jt})_t$ describing acceleration cycles in countries i and j, respectively. In this chapter, we test the hypothesis that acceleration cycles are either strongly non-synchronized [SNS] or strongly perfectly positively synchronized [SPPS] based on the statistic $\hat{\rho}_S$, namely the estimated correlation coefficient between $(S_{it})_t$ and $(S_{jt})_t$. Harding and Pagan (2006) establish a relationship between the estimated concordance index $\widehat{CI}$ and correlation coefficient $\hat{\rho}_S$, showing that:

$$\widehat{CI} = 1 + 2\hat{\sigma}_S + 2\hat{\mu}_{S_i}\hat{\mu}_{S_j} - \hat{\mu}_{S_i} - \hat{\mu}_{S_j}, \tag{14.9}$$

where $\hat{\mu}_{S_i} = E(S_{it})$, $\hat{\mu}_{S_j} = E(S_{jt})$ and $\hat{\sigma}_S$ is the covariance between $(S_{it})_t$ and $(S_{jt})_t$ such that:

$$\hat{\sigma}_S = \hat{\rho}_S \sqrt{\hat{\mu}_{S_i}(1 - \hat{\mu}_{S_i})} \sqrt{\hat{\mu}_{S_j}(1 - \hat{\mu}_{S_j})}. \tag{14.10}$$

First, as suggested in Harding and Pagan (2006), the null SNS hypothesis $\rho_S = 0$ can be tested starting from the following regression equation:

$$\hat{\sigma}_{S_i}^{-1}\hat{\sigma}_{S_j}^{-1}S_{jt} = a + \rho_S\hat{\sigma}_{S_i}^{-1}\hat{\sigma}_{S_j}^{-1}S_{it} + u_t, \tag{14.11}$$

where $\hat{\sigma}_{S_i}^2$ and $\hat{\sigma}_{S_j}^2$ are the estimated variances of $(S_{it})_t$ and $(S_{jt})_t$, respectively. In business cycle analysis, both variables $(S_{it})_t$ and $(S_{jt})_t$ involved in the previous regression equation, often present strong auto-correlation due to the duration of cycle phases. For example, for the euro area as a whole, the auto-correlation function for the first lag is equal to 0.84 for IPI and to 0.53 for GDP. Thus, testing the null $\rho_S = 0$ requires to take auto-correlation, as well as heteroscedasticity of the errors $(u_t)_t$, into account using standard procedures. In this respect we use a heteroscedastic and auto-correlation consistent (HACC) standard error with Bartlett weights, the number of lags being suggested by Newey and West (1984).

In addition to the previous bivariate tests, the multivariate test of Harding and Pagan (2006) enables to test the null hypothesis of strong multivariate non-synchronization (SMNS) among n countries. We recall briefly the test procedure and we refer to the original paper for further details. This GMM-based procedure starts from the following $n(n + 1)/2$ moment conditions:

$$E(h_t(\theta, S_t)) = 0, \tag{14.12}$$

where

$$h_t = \begin{bmatrix} S_{1t} - \mu_{S_1} \\ \vdots \\ S_{nt} - \mu_{S_n} \\ \frac{(S_{1t}-\mu_{S_1})(S_{2t}-\mu_{S_2})}{\sqrt{\mu_{S_1}(1-\mu_{S_1})\mu_{S_2}(1-\mu_{S_2})}} - \rho_S^{12} \\ \vdots \\ \frac{(S_{(n-1)t}-\mu_{S_{(n-1)}})(S_{nt}-\mu_{S_n})}{\sqrt{\mu_{S_{(n-1)}}(1-\mu_{S_{(n-1)}})\mu_{S_n}(1-\mu_{S_n})}} - \rho_S^{(n-1)n} \end{bmatrix},$$

and where $\theta' = (\mu_1, \ldots, \mu_n, \rho_S^{12}, \ldots, \rho_S^{(n-1)n})$. For the SMNS case, the restricted parameter vector of dimension $n(n+1)/2$ is such that $\theta_0' = (\mu_1, \ldots, \mu_n, 0, \ldots, 0)$. Under the null of SMNS, the statistic

$$W_{SMNS} = \sqrt{T}g(\theta_0, \{S\}_{t=1}^T)'\hat{V}^{-1}\sqrt{T}g(\theta_0, \{S\}_{t=1}^T) \tag{14.13}$$

where $g(\theta_0, \{S\}_{t=1}^T) = \frac{1}{T}\sum_{t=1}^T h_t(\theta_0, S_t)$ and where $\hat{V}$ is the robust HACC estimator of the variance-covariance matrix of $\sqrt{T}g(\theta_0, \{S\}_{t=1}^T)$, has an asymptotic $\chi^2_{(n(n-1)/2)}$ distribution.

An application of those concepts can be found in Darné and Ferrara (2011), as regards the euro area acceleration cycle.

14.3.3 Comparing chronologies

In this subsection, we show how chronologies can vary when alternative dating algorithms are applied or when we refer to different variables. As a first example the chronology from Bruno and Otranto (2008) is proposed. The historical results obtained with the different methods are compared with the dating provided by ISAE.

In order to show how the chronologies might look like depending on the method used, the table below summarizes six different methods used to achieve the same goal - namely the chronology for the Italian economy. Table 14.1 shows the Italian chronology based on six different dating methods.

The names of the different methods are as follows:

ISAE: official dating provided by ISAE;

INDNP: Indirect non-parametric detection;

INDMIX: Indirect mixed detection;

DIRNP: Direct non-parametric detection;

DIRMIX: Direct mixed detection;

DIRP: Direct fully parametric detection.

The results obtained show a tendency of the direct methods to find out more cycles than those detected by ISAE (three more for DIRNP and DIRMIX, two more for DIRP), while the indirect methods are more reliable with respect to this point, having detected only one extra-cycle each.

The second example is based on Darné and Ferrara (2009) and shows chronologies for acceleration cycles for the six major European economies. Table 14.2 shows cycles based on GDP, table 14.3 the ones based on IPI.

Table 14.1: Italian chronology based on six different dating methods

Turning points	ISAE	INDNP	INDMIX	DIRNP	DIRMIX	DIRP
Peak	I 74	XII 73	II 74	II 74	I 74	I 74
Trough	V 75	VIII 75	V 75	VIII 75	VIII 75	V 75
Peak	II 77	XII 76	XII 76	I 77	XII 76	XI 76
Trough	XII 77	I 78	I 78	XII 77	XII 77	IX 77
Peak	III 80	I 80	II 80	I 80	III 80	XI 79
Trough				VI 81		
Peak				II 82		
Trough	III 83	III 83	II 83	II 83	V 83	V 83
Peak			VIII 85		VIII 84	VIII 85
Trough			I 86		I 85	I 86
Peak		VIII 89		IV 90	XI 89	II 90
Trough		VI 90		XII 90	III 91	
Peak	III 92	II 92	XII 91	I 92	II 92	
Trough	VII 93	VII 93	VII 93	VIII 93	VIII 93	VIII 93
Peak	XI 95	X 95	VIII 95	XII 95	VIII 95	XII 95
Trough	XI 96	VIII 96	IX 96	VI 96	XII 96	XII 96
Peak				VII 98	XII 97	XII 97
Trough				I 99	XII 98	IV 99
Peak	XII 00	X 00	X 00	XI 00	XII 00	XI 00
Trough		XII 01	XII 01	III 02		

14.3.4 Euro area chronologies

Since 2007, Eurostat started to elaborate and update on quarterly basis euro area chronologies for the growth and the business cycles according to the ABCD approach. Starting from 2011, also the acceleration chronology was added. Together with the euro area, chronologies for 11 economies have been regularly compiled and updated: Germany, France, Italy, Belgium, the Netherlands, Spain, Portugal, Austria, Finland, Greece and Ireland. The compilation of the chronologies for the remaining countries is still ongoing. The euro area final chronology is obtained by comparing the direct dating based on euro area aggregates with an indirect one obtained by a weighted average of national chronologies. A particular effort in compiling the euro area chronologies is made in order to minimise the discrepancies between the direct and indirect dating. The dating methodology used is the one described in 3.2. Table 14.4 shows the real turning points identified for the euro area acceleration, growth and business cycles, starting in 1973.

* provisional dating

We already mentioned the importance of keeping past turning points constant across different releases of the chronologies. Nevertheless, since in the latest years, variables unavoidably subject to revision, the chronologies related to the recent time periods are labelled as provisional. As a general rule we consider that statistical variables can be revised over a maximum 3-4 years so that during this periods we accept that the turning points can change across different releases. By contrast, past turning points are frozen and, only in some exceptional cases, we accept to revise them.

14.4 Construction of turning point indicators

In this section, we review the various steps towards the construction of turning point indicators for each type of economic cycle. We focus our attention on variable selection and model selection aspects, as well as on

Table 14.2: Acceleration cycles based on GDP

Turning point	Euro	Germany	France	Italy	Spain	Belgium	Netherlands
Trough	1996 Q1	1996 Q1	1996 Q1	1996 Q2		1996 Q1	1996 Q1
Peak	1997 Q2	1997 Q2	1998 Q2	1997 Q2	1997 Q4	1997 Q1	1997 Q2
Trough	1998 Q4	1998 Q2	1998 Q4	1998 Q4	1998 Q4	1998 Q3	1998 Q2
Peak	1999 Q3	1999 Q4	1999 Q4	1999 Q4	1999 Q2	1999 Q3	1999 Q1
Trough	2001 Q3	2000 Q3	2001 Q2	2001 Q2	2002 Q1	2001 Q3	
Peak		2001 Q1					
Trough		2002 Q1					2002 Q1
Peak	2002 Q2	2002 Q3	2002 Q1	2002 Q2	2003 Q1	2002 Q3	
Trough	2003 Q2	2003 Q1	2002 Q4	2003 Q1	2003 Q3	2003 Q1	
Peak	2004 Q1	2003 Q3	2004 Q2	2004 Q1	2004 Q3	2004 Q2	2004 Q1
Trough	2004 Q4	2004 Q3	2005 Q2	2004 Q4		2005 Q1	2005 Q1
Peak	2006 Q2	2006 Q2	2006 Q2	2006 Q4		2005 Q4	

Table 14.3: Acceleration cycles based on IPI

Turning point	Euro	Germany	France	Italy	Spain	Belgium	Netherlands
Peak		XII 1992	X 1991	VIII 1991	IX 1991	XI 1991	
Trough	XI 1992	XII 1992	XI 1992	VII 1992	XI 1992	XI 1992	
Peak	III 1994	IX 1994	IV 1994	III 1994	VI 1994	I 1995	
Trough	II 1996	X 1995	IX 1995	III 1996	I 1996	IX 1995	III 1995
Peak	IV 1997	XI 1997	IV 1997	IV 1997	III 1997	VI 1997	XI 1995
Trough	X 1998	IX 1998	VIII 1998	XI 1998	II 1999	X 1998	I 1997
Peak	IX 1999	IV 2000	IX 1999	VIII 1999	XI 1999	VIII 1999	I 2000
Trough	IX 2001	IX 2001	X 2001	VII 2001	X 2001	IV 2001	VIII 2001
Peak	IV 2002	VI 2002	IV 2002	IV 2002	VIII 2002	III 2002	IV 2002
Trough	IV 2003	IV 2003	IV 2003	II 2003	V 2003	XI 2002	IV 2003
Peak	X 2003	XI 2003	X 2003	VIII 2003	II 2004	II 2004	I 2004
Trough	XI 2004	X 2004	III 2005	XI 2004	X 2004	XII 2004	XII 2004
Peak	IV 2006	V 2006	XI 2005	I 2006	V 2006	XII 2005	XII 2005
Trough		III 2007	VIII 2006				VII 2006
Peak			II 2007				

the criteria for evaluating the indicators. The section ends with an illustration of this approach based on the euro area coincident indicators, currently compiled by Eurostat.

14.4.1 Data selection

Generally, datasets used for the development of cyclical indicators are stemming from three main sources of information: macroeconomic data (hard data), opinion surveys (soft data) and financial data. Hard data is strongly correlated with business cycles but is well known for its lack of timeliness: it is indeed published with a strong delay and is often revised from one month to the other. This constitutes a major drawback for practitioners involved in real-time analysis. Financial variables have been proved to be leading towards the global economic cycle in many empirical studies and are consequently rather introduced in leading indicators of the cycle (see among others Estrella and Mishkin (1998), or Stock and Watson (2002)). Especially, term spread, namely the difference between long-term and short term interest rates, has proved to be a leading index of recessions (see for example, Estrella et al. (2003), Kauppi and Saikkonen (2008), or Rudebusch and Williams (2009)). Also stock prices (Farmer, 2011) or oil prices (Hamilton (2003)) are leading variables that can be integrated in models to anticipate turning points. Bellégo and Ferrara (2012) have also shown that

Table 14.4: Chronology for the Euro Area - $\alpha AB\beta CD$ sequences

Period	α	A	B	β	C	D	Cycle type
1973-74	1973 Q1	1974 Q1	1974 Q2	1974 Q3	1975 Q1	1975 Q3	Pure Business Cycle
1975-77	1975 Q4	1977 Q1		1977 Q2		1978 Q2	Pure Growth Cycle
1978-81	1978 Q4	1979 Q4	1980 Q1	1980 Q2	1980 Q4	1981 Q1	Pure Business Cycle
1981-82	1981 Q1	1981 Q4	1981 Q4	1982 Q2	1982 Q4	1982 Q4	Pure Business Cycle
1983-84	1983 Q4			1984 Q2			Pure Acceleration Cycle
1985-87	1985 Q4	1986 Q1		1986 Q4		1987 Q2	Pure Growth Cycle
1987-89	1987 Q4			1989 Q2			Pure Acceleration Cycle
1991-93	1991 Q1	1992 Q1	1992 Q1	1992 Q3	1993 Q3	1993 Q3	Pure Business Cycle
1995-96	1995 Q1	1995 Q1		1996 Q1		1996 Q4	Pure Growth Cycle
1997-98	1997 Q2	1998 Q1		1998 Q4		1999 Q1	Pure Growth Cycle
2000-01	2000 Q1	2000 Q3		2001 Q3			Double dip inside Growth Cycle
2002-03	2002 Q2			2003 Q1		2003 Q3	Double dip inside Growth Cycle
2003-05	2003 Q4			2005 Q1			Pure Acceleration Cycle
2006-09	2006 Q2	2008 Q1	2008 Q1	2009 Q1	2009 Q2	2009 Q3	Pure Business Cycle
2010-13	2010 Q2	2011 Q3*	2011 Q3*	2012 Q4*	2013 Q1*	2013 Q1*	Pure Business Cycle

aggregating financial information from several variables through a factor-probit model can be a fruitful strategy when one aims at anticipating cyclical turning points. Opinion surveys are the most frequently watched variables in economic institutions. They convey useful information as regards the current state of the economic cycle. They also possess the great advantage of not being revised after their release and are timely available, usually before the end of the reference month.

Choosing among a large set of economic variables is not an easy task. Generally two approaches are taken. The first one consists in reducing the dimension of the problem by estimating a dynamic factor model as the ones proposed by Stock and Watson (2002) or Forni et al. (2005). The estimated factors are a linear combination of original variables and take into account the cross-correlation among those variables. Estimated factor can in turn be used as coincident or leading indicators of economic cycles. The second approach relies on a small set of variables (carefully selected for their ability to track economic cycles) from a larger dataset. The selection process can be done by optimizing a given criterion, such as the Quadratic Probabilistic Score (QPS, see Diebold and Rudebusch (1989)) for example, or by using a selection algorithm such as the LASSO one. In this chapter, we privilege this second approach based on a narrow set of variables keeping in mind that it is easier to explain changes in the values of the indicator in real-time, by comparison with information extracted from a large dataset.

14.4.2 Model selection

Starting from a n-vector of variables $(x_t^1, \ldots, x_t^n)$, the objective is to compute an indicator lying between 0 and 1 such that the economy belongs to the low phase of the cycle when the indicator is close to one and belongs to the high phase of the cycle when the indicator is close to zero. In this respect, non-linear models that provide a probability of being in a given regime at any given date t are of great interest. A review of such models is proposed by Ferrara and Mazzi (2016). Among those non-linear models, it turns out that multivariate Markov-Switching models are of great interest and have proved their ability to reproduce business cycles; we refer among others to Hamilton (1989), Krolzig (1997), Chauvet (1998), Kim and Nelson (1998), Ferrara (2003), Anas and Ferrara (2004b), Chauvet and Hamilton (2006), Bengoechea et al. (2006), Anas et al. (2008), Camacho and Perez-Quiros (2010), Darné and Ferrara (2011).

Multivariate models

Markov-Switching models were first introduced in the business cycle literature by Hamilton (1989) to deal with non-linear time-series. In particular, Markov-Switching models were originally proposed to model discrete shifts in the mean growth rate of a non-stationary time-series, that is, episodes across time in which the dynamic behavior of a series undergoes abrupt changes. The first application of this class of models was to the U.S. business cycle.

Following the literature on VAR models, Hamilton (1989) used an auto-regressive process to approximate an observable non-stationary, in the sense stated above, time-series, namely the U.S. real GNP. Hamilton's seminal idea was to assume the parameters of the auto-regressive model to be time-varying and evolving according to a latent Markov-chain process. Conditional on the unobservable variable, the auto-regressive model is assumed to be time-invariant. In this respect, Markov-Switching models are an extension of traditional VAR models.

As for the latent state-variable, it is only natural to assume that regime changes are not directly observed by the researcher; instead, he or she must draw inference on their occurrence based on the realizations of the observable time-series. The estimated probability of occurrence of a shift in the regime of the latent variable is used, within our investigation on economic cycles, to assess the prevailing economic regime (being, say, either contraction or expansion) at any given point in time.

Formally, the most general specification of the Markov-Switching model we consider is one in which all the parameters of the auto-regressive model are conditional on the state of the latent Markov-chain (s_t):

$$y_t = \alpha(s_t) + \sum_{j=1}^{p} \beta_j(s_t)y_{t-j} + \epsilon_t, \tag{14.1}$$

where $\epsilon_t \ N(0, \sigma^2(s_t)), t = 1, \ldots, T$. In Hamilton's original model the observable endogenous variable y_t is the quarterly percentage change in U.S. real GNP, so that the observable variable is piecewise stationary[3].

The definition of the data generating process requires, in addition to the model for the observable time-series, the specification of the process followed by the latent variable. As stated above, a first-order ergodic discrete-state Markov-chain is the stochastic process that governs the realization of the unobservable state-variable. Under the assumption of time-invariant transition probabilities, the Markov-chain process above is defined by the transition probabilities in (14.2):

$$Pr(s_t = j|s_{t-1} = i, s_{t-2}, s_{t-3}, \ldots) = Pr(s_t = j|s_{t-1} = i) = p_{i,j}, \tag{14.2}$$

for $i, j = 1, \ldots, M$. Transition probabilities measure the probability of either staying in the same regime or switching to another regime in moving from time $t - 1$ to time t. The definition of a first-order Markov-chain implies that the probability of observing regime j at time t depends only on the regime prevailing at the previous time.

Hamilton (1990) also proposed a non-linear iterative algorithm named Expectations Maximization (EM) that allows estimating the population parameters (auto-regressive coefficients and transition probabilities) by Maximum-Likelihood.

The filtering and smoothing algorithms that are embedded in the EM technique allow, as a by-product of the estimation process, to draw inference on the probability of the prevailing regime of the latent variable.

---

[3]It is worth noting that in model (14.1) the intercept is state-dependent, whereas Hamilton (1989) considered a mean-adjusted form. Contrary to linear VAR model, these two specifications are not equivalent for the class of Markov-Switching VAR models. We refer to Krolzig (1997) for a proof.

The two algorithms above are actually named after the estimate they provide of the smoothed and filtered probabilities, respectively, of the unobserved states of the Markov-chain. Smoothed probabilities are defined as the estimated probabilities of observing a state of the latent Markov-chain at time t given the whole sample information of the observed time-series. Filtered probabilities differ from them since they are conditional only on the observed variable available at time t. For a thorough description of the EM estimation method we refer to Hamilton (1990) and Krolzig (1997).

Despite several extensions of Hamilton's original model proposed in the literature, in our application to the estimation of probabilistic coincident indicators of the Euro area's economic cycles we prefer to focus on parsimonious Markov-Switching models. As a matter of fact, we empirically found more convenient to assume that only some parameters are dependent on the state-variable whereas the remaining ones are regime invariant. More in detail, we used model specifications in which the intercept term and the variance of the error term depend on the same latent state-variable. Furthermore, the order of the auto-regressive polynomial is set at zero. Therefore, although in a trivial way, the auto-regressive coefficients are time-invariant and not subject to changes in regime.

Following a notation firstly introduced by Krolzig (1997), the models we used in our empirical research can be denoted as MSIH(K)-VAR(0), where MS obviously stands for Markov-Switching; the letters I and H indicate, respectively, that the intercept and variance of the errors are time-dependent as they are governed by a common latent discrete Markov-chain. The number of regimes of this latent variable is K. Finally, VAR(0) indicates that the order of the lag polynomial is zero. Obviously, a particular case of this representation, when the vector of variables X is unidimensional, is the MS model of the form MSIH(K)-AR(0).

As for the estimation of the model parameters, we applied the EM method referred to above as it was coded in the MSVAR algorithm by Krolzig.

Constructing composite indicators for turning point detection

In the context of euro area's economic cycles, we aim at constructing composite probabilistic indicators to detect in real-time the occurrence of a shift in economic regime of the business, growth and acceleration cycles. As it is our purpose to provide real-time signals, the relevant measure of the prevailing regime is the filtered probabilities. We thereby use Markov-Switching models to estimate filtered probabilities from several economic variables and then aggregate them into probabilistic composite indicators.

The construction of composite probabilistic indicators heavily relies on the interpretation of the Markov-Switching models in terms of economic cycle. This is achieved by assigning an economic meaning to the latent state-variable: its regimes can be thought of representing the different phases of the economic cycle. For example, in the case of a two-state Markov-chain, it is straightforward to interpret one regime (say, the one with a positive intercept) as expansionary and the other one (say, the one with a negative intercept) as recessionary. From the association between states of the latent variable and economic regimes follows immediately that filtered probabilities can be interpreted as a measure of the probability of observing a given phase of the cycle at every point in time. This interpretation of the filtered probabilities in terms of economic cycle paves the way to the construction of the composite indicators, which is the goal of our investigation.

As we aim at constructing composite coincident indicators of the Euro area's business, growth and acceleration cycles, we have to define what regimes of the latent Markov-chain are associated with recession, slowdown and deceleration periods, respectively. Such an issue could in principle be settled a priori if the researcher has strong believes, perhaps based on an economic theory, about what regimes of the latent state-variable are actually representing specific phases of the economic cycle. In our empirical analysis we rather preferred to follow a more flexible approach and relate states of the latent variable to economic regimes by looking ex-post at the signs and magnitudes of the state-dependent intercept.

In the following section, for each of the component variables of the coincident indicators we built, we will provide a description of the interpretation we gave to the states of the latent Markov-chain.

The final point we make concerns to two issues that so far have remained in the background, that is, how to select the component variables and how to aggregate them. As far as the variable selection is concerned, we focus on variables included in the PEEIs (Principal European Economic Indicators). They span from hard economic variables (e.g. industrial production and unemployment rate) to soft variables (e.g. business surveys). As we will show in a later section, the first class of variables plays a major role in obtaining a coincident indicator of the business cycle, whereas soft variables are mainly used for the acceleration cycle. The growth cycle coincident indicator is obtained by using both hard and soft data.

As for the aggregation method, we used two different approaches that are tied to two different specifications of the Markov-Switching models. In the first one, a single equation Markov-Switching model is separately fitted to each component variable and the filtered probabilities that are obtained as a by-product are aggregated and the composite coincident indicator is finally constructed as a weighted average:

$$\text{Indicator}_t = \sum_{k=1}^{N} w_k Pr(\text{Downturn}_t^k),\qquad(14.3)$$

where $Pr(\text{Downturn}_t^k)$ is the probability that the $k - th$ component variable is either in a recession, slowdown or deceleration of the respective cycle at time t. w_k is the weight given to the $k - th$ component; the weight attached to the downturn probabilities of each component variable reflects the accuracy in locating the phases of the economic cycle under study. The dating chronologies presented in a previous section are used as benchmark for this assessment. In other words, following this approach, the coincident indicators for the business, growth and acceleration cycles are independently compiled by using the most appropriate set of variables and weighting scheme. The main drawbacks of this approach are the role that subjective appreciation can play in the definition of the weighting scheme and the fact that there is no guarantee that the three indicators correctly follow the $\alpha AB\beta CD$ sequence.

The second approach entails the estimation of a multivariate Markov-Switching model, instead of the several single equation models above. It immediately follows that the aggregation procedure in the second case is embedded in the estimation of the multivariate model. In this respect, the researcher loses control of the aggregation step. If, on the one hand, this avoids his or her discretion in defining the weighting scheme, on the other hand does impede to build in further knowledge into the composite indicator, such as for example the accuracy of each component in locating downturn phases of the economic cycle.This approach can obviously allow for the independent estimation of the coincident indicators for the business, growth and acceleration cycles, based on multivariate models. Nevertheless, the biggest advantage of following this approach is represented by the possibility of jointly estimating coincident indicators for the growth and business cycles in a multivariate context. To achieve this objective, variables able to explain both cycles have to be selected and, within the MS-VAR model, the number of regimes has to be increased in order to describe all economic phases. The jointly construction of the business and growth cycle indicators will avoid any risk inconsistency with the ABCD sequence. Unfortunately, the multivariate modelling strategy cannot be extended also to the acceleration cycle due to the asymmetry characterising the sequence of peaks with respect to the sequence of troughs within the $\alpha AB\beta CD$ approach.

14.4.3 Assessment of turning point indicators

To assess the quality of the model concerning business cycle replication (in-sample evaluation) or concerning the forecast evaluation (out-of-sample evaluation) several criteria have been proposed. Probability forecasts are often evaluated using the quadratic probabilistic score of Brier (1950) defined by :

$$QPS = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{P}_t - r_t \right)^2, \tag{14.4}$$

where $\hat{P}_t$ is the estimated probability of being in a given regime (e.g. recession) and r_t is the reference one-zero variable. The QPS is bounded between 0 and 1, values close to zero implying a good quality in terms of business cycle replication, and conversely.

Another statistic computed as goodness-of-fit measure to the reference chronology is the Concordance Index, which is defined as follows:

$$CI = \frac{1}{T} \left[\sum_{t=1}^{T} I_t \times r_t + \sum_{t=1}^{T} (1 - I_t) \times (1 - r_t) \right], \tag{14.5}$$

where r_t is the reference binary variable already used in the QPS and where I_t is a binary random variable that takes value 1 if $\hat{P}_t > \kappa$ and 0 otherwise, where κ is a given threshold between 0 and 1. Following the approach proposed by Hamilton (1989), the natural critical value of $\kappa = 0.5$ is often chosen. However, the choice of κ can be discussed and selected by ad-hoc methods according to a specific criterion. Obviously, there is a trade-off between the timeliness and the reliability of the signal. Other threshold values κ different from 0.5 may be used to provide a better goodness of fit to the reference turning point chronology. This point is rarely considered in research papers but has major practical implications as regards the economic interpretation of the signal. A good practice would be to assess the sensitivity of the signal to the values of κ as for example in Darné and Ferrara (2011).

Another often used measure is the log-probability score (LPS) defined by:

$$LPS = \frac{1}{T} \sum_{t=1}^{T} \left[r_t \log(\hat{P}_t) + (1 - r_t) \log(1 - \hat{P}_t) \right]. \tag{14.6}$$

The LPS is non-negative and penalizes large mistakes more heavily than the QPS. The Kuipers Score is also sometimes used and is defined by:

$$KS = H - F, \tag{14.7}$$

where H is the hit rate, that is the proportion of total number of regimes 1 that were correctly reproduced, and where F is the false signal rate, that is the proportion of the total number of regimes 0 that were incorrectly estimated as being regime 1. Those rates can be estimated respectively by :

$$H = \frac{\sum_{t=1}^{T} r_t I_t}{\sum_{t=1}^{T} r_t} \tag{14.8}$$

and

$$F = \frac{\sum_{t=1}^{T} (1 - r_t) I_t}{\sum_{t=1}^{T} (1 - r_t)}, \tag{14.9}$$

where I_t is defined above. We refer also to Chauvet and Hamilton (2006) for the concept of hitting probabilities.

14.4.4 A set of euro area coincident indicators: An empirical illustration

Since 2007, Eurostat started the compilation of a set of euro area coincident indicators for a real-time monitoring of the euro area cyclical situation. Those indicators are regularly compiled on a monthly basis and they refer to acceleration, business and growth cycle. The three indicators are labelled Acceleration Cycle Coincident Indicator (ACCI), Business Cycle Coincident Indicator (BCCI) and Growth Cycle Coincident Indicator (GCCI). Since 2011, Eurostat also started to compile a pair of coincident indicators for growth cycle and business cycle based on a multivariate Markov-Switching model labelled respectively MS-VAR GCCI and MS-VAR BCCI. They aimed to overcome the risk of inconsistencies between the BCCI and the GCCI due to the different lags in detecting turning points.This subsection describes those indicators and shows their behaviour based on the latest releases.

ACCI

The coincident indicator of the acceleration cycle (ACCI) returns the probability of a deceleration of the euro area growth rate cycle. As stated above, this is not really a composite indicator as it is based upon only one variable, namely the economic sentiment indicator (ESI). More specifically, the ACCI is obtained by fitting a MSI(3)-AR(0) model to the ESI differenced twice, first over 6 months and then over 1 month. The double differentiation of the endogenous variable is consistent with the definition of acceleration as change in the pace of growth.

As for the interpretation of the above model in terms of the growth rate cycle, the first regime, which is the only regime for which the state-dependent intercept takes on a negative value, is assumed to identify the deceleration phases of the growth rate cycle. Deceleration probabilities are accordingly set equal to the filtered probabilities of this regime.

The latest release of the ACCI estimated in June 2015 is graphically illustrated in Figure 14.4 below.

Over the period stretching from August 1985 to December 2014, the ACCI does not miss any of the ten decelerations of the euro area acceleration cycle (type-I error). However, it signals a deceleration between mid-1989 and early 1991 that was not identified in the reference dating chronology (type-II error). Further, between January 2014 and January 2015 the ACCI has been signaling a deceleration not identified in the reference chronology; however, as the reference chronology over the last two years is still provisional, evidence for a false deceleration is not yet conclusive. Overall, the ACCI is quite accurate in signaling the decelerations of the growth rate cycle. Four out of the ten peaks are detected with an average delay of 0.9 months, whereas other four peaks are signaled 1.7 months in advance, on average. The remaining two peaks are exactly located as in the reference chronology.

The accuracy of the ACCI is summarized by the QPS and Concordance Index statistics reported in Table 14.1.

Table 14.1: ACCI's Accuracy Statistics

QPS	Concordance Index
0.22	0.71

Figure 14.4: ACCI from August 1985 to May 2015 (blue line) and decelerations of the growth rate cycle (grey shaded areas).
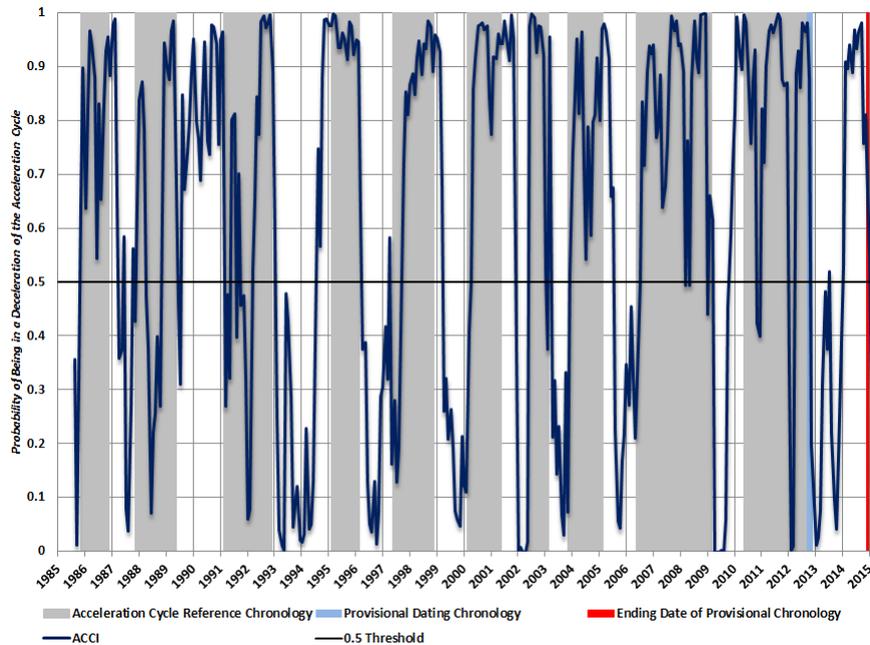
GCCI

The Growth Cycle Coincident Indicator (GCCI) is a measure of the probability of slowdown of the euro's area growth cycle. The GCCI is obtained as an equally weighted average of the slowdown probabilities estimated from five component variables. The first two components pertain to the real economy, namely, the industrial production index and the imports of intermediate goods from outside the euro area. The remaining three components are all surveys, namely, the employment expectations in the industry for the months ahead, the construction confidence indicator and the consumer confidence indicator.

All the five GCCI's components are seasonally adjusted. The first two components, namely the ones related to the real economy, are differenced twice; only one differentiation is taken for each of the three soft variables. As shown in the Table below, a MSIH(4)-AR(0) model is fitted to the IPI and Imports variables. The slowdown phases are obtained by summing the filtered probabilities of the first two regimes, as these two regimes are characterized by negative values of the state-dependent intercept. A MSI(2)-AR(0) model is fitted to the survey variables, such that the slowdown periods are associated to the first regime, that is, to the state for which the intercept takes on a negative value.

Table 14.2: GCCI's Components

Variable	Seasonal Adjustment	Source	Sample	Differentiation	MS Model	Recession Regime(s)	Weight
Industrial Production Index	SA	Eurostat	1991-2015	12- and 6-month	MSIH(4)-AR(0)	R1+R2	0.20
Imports of Intermediate Goods	SA	Eurostat	1991-2015	12- and 6-month	MSIH(5)-AR(0)	R1+R2	0.20
Employment Expectations	SA	DG-EcFIN	1985-2015	6-month	MSI(2)-AR(0)	R1	0.20
Construction Confidence Indicator	SA	DG-EcFIN	1985-2015	6-month	MSI(2)-AR(0)	R1	0.20
Consumer Confidence Indicator	SA	DG-EcFIN	1985-2015	6-month	MSI(2)-AR(0)	R1	0.20

The GCCI is compared to the reference dating chronology of the growth cycle over the period July 1991 - December 2014 to assess the accuracy of the former in locating the slowdowns experienced in the euro area. The GCCI signals all the six slowdowns identified in the reference chronology (no type-I error). However, the slowdown caused by the Asian crisis between 1998 and 1999 is detected by the GCCI with some delay (7 months) and not as strongly as the other five slowdowns (the GCCI does not exceed 0.6 during this period).

No false slowdowns (type-II error) are signaled by the GCCI; however, the start of the 2008-2009 slowdown is detected 11 months in advance compared to the reference chronology. This is an exception, as three out of the five peaks in the reference chronology are signaled with an average delay of 2.6 months. Recently, the GCCI had been increasing in the second half of 2014, though it remained below the 0.5 threshold. This increase was entirely due to the slowdown probabilities derived from the IPI and Imports components.

Figure 14.5: GCCI from July 1991 to April 2015 (blue line) and slowdowns of the growth cycle (grey shaded areas).
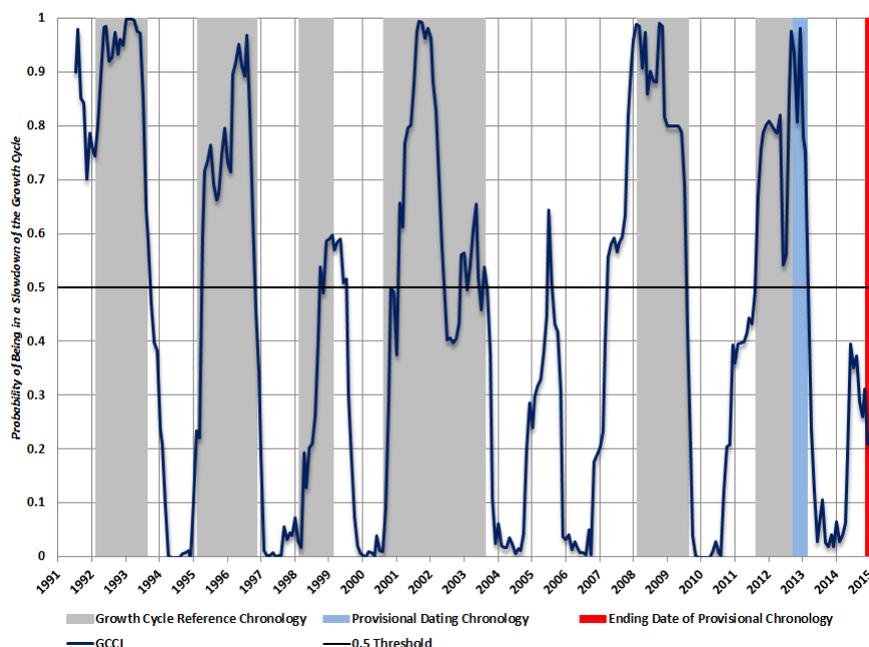
Accuracy statistics (QPS and Concordance Index) for the GCCI are reported in Table 14.3.

Table 14.3: GCCI's Accuracy Statistics

QPS	Concordance Index
0.13	0.84

BCCI

The probability of a recession of the euro area business cycle is provided by the BCCI. Three variables are included in the BCCI: Industrial Production Index, unemployment rate and new passenger car registrations. All the three components of the BCCI are variables pertaining to the real economy and seasonally adjusted.

Following Hamilton (1989), the endogenous observable variables are differenced to achieve stationarity. The order of differentiation and the MS model for each variable is chosen empirically case by case, with the aim of reaching the highest accuracy in locating the recessions of the business cycle. For all the component variables, the regime of the latent variable that is assumed to identify recession phases of the business cycle is the first one, that is, the regime for which the state-dependent intercept is the lowest. This is a negative value for all the three models. However, for the IPI and unemployment rate models, also the second regime is characterized by statistically significant negative intercepts, though this regime is not assumed to identify recession periods.

Finally, the composite probabilistic indicator of the business cycle is obtained as a weighted average of the recession probabilities obtained from the three components. The weight given to each variable is proportional

to the number of phases of the business cycle correctly located and inversely related to the number of phases of the business cycle that are missed or falsely detected.

Table 14.4: BCCI's Components

Variable	Seasonal Adjustment	Source	Sample	Differentiation	MS Model	Recession Regime(s)	Weight
Industrial Production Index	SA	Eurostat	1971-2015	12-month	MSIH(4)-AR(0)	R1	0.34
Unemployment Rate	SA	Eurostat	1976-2015	3-month	MSI(3)-AR(0)	R1	0.46
New Passenger Car Registrations	SA	ACEA	1979-2015	12-month and MA3	MSI(3)-AR(0)	R1	0.20

When comparing the BCCI with the reference dating chronology of the business cycle it turns out that the BCCI locates all the five recessions suffered by the euro area between 1979 and 2014. The main backdrop of the BCCI is its lag in detecting peaks and troughs, 7.0 and 5.6 months on average, respectively.

Figure 14.6: BCCI from June 1979 to April 2015 (blue line) and recessions of the business cycle (grey shaded areas).
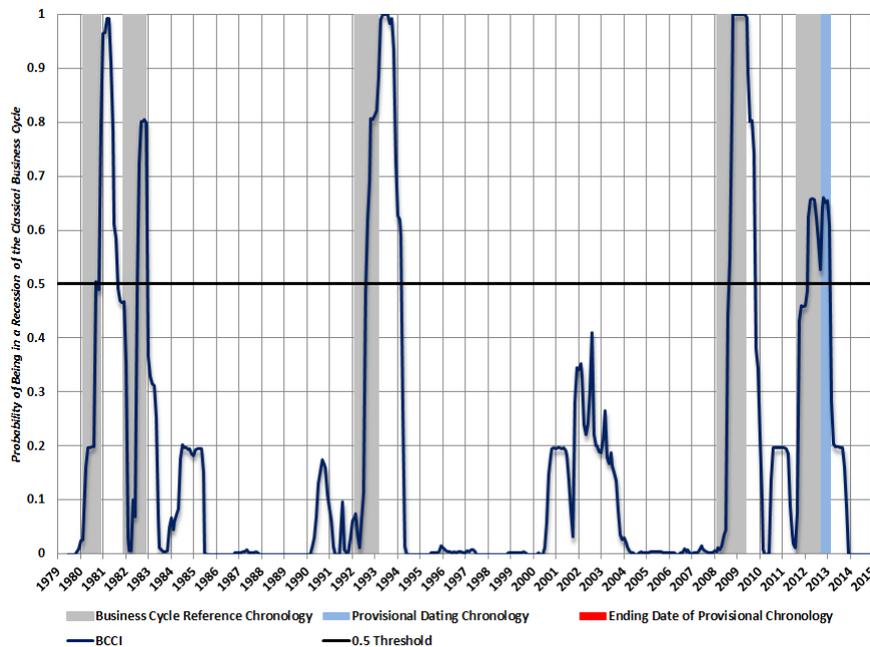
QPS and Concordance Index statistics related to the BCCI are reported in Table 14.5.

Table 14.5: BCCI's Accuracy Statistics

QPS	Concordance Index
0.12	0.85

MS-VAR GCCI and MS-VAR BCCI

It should be noted that the BCCI and GCCI are obtained separately one from another as single-equation Markov-Switching models are estimated independently and the recession/slowdown probabilities estimated from them are then aggregated. The downside is that turning points signaled by these two probabilistic indicators need not to comply with the ABCD approach. Indeed, in the overlapping period 1991-2014 of these two probabilistic indicators, the trough (C) of two out of the three recessions are detected after the

corresponding trough (D) of the growth cycle: March 1994 (C) vs. September 1993 (D) and October 2009 (C) vs. July 2009 (D).

To overcome this drawback we devised the estimation of coincident indicators of the business and growth cycles that were consistent by construction with the ABCD approach. We built this pair of coincident indicators through a multivariate Markov-Switching model. We empirically specified a MSIH(4)-VAR(0) and fitted it to four variables over the period February 1985 - April 2015: the industrial production index, the unemployment rate, new passenger car registrations and the employment expectations in the industry.

The filtered probabilities obtained as a by-product of the estimation of the Markov-Switching model above are used to jointly construct both a coincident indicator of the business cycle (MS-VAR BCCI) and a coincident indicator of the growth cycle (MS-VAR GCCI).

The first regime of the latent Markov-chain is assumed to correspond to periods of recession, thereby the MS-VAR BCCI is set equal to the filtered probabilities estimated for this regime. The co-joint coincident indicator of the growth cycle (MS-VAR GCCI) is obtained by summing the filtered probabilities of the first two regimes.

The endogenous variables of the model, as well as their order of differentiation considered in the model specification, are summarized in Table 14.6.

Table 14.6: Components of the MS-VAR GCCI and MS-VAR

Variable	Seasonal Adjustment	Source	Differentiation Order
Industrial Production Index	SA	Eurostat	12-month
Unemployment Rate	SA	Eurostat	1-month
New Passenger Car Registrations	SA	ACEA	3-month
Employment Expectations	SA	DG-EcFIN	1-month

It is worth noticing that, contrary to the BCCI and GCCI, no explicit weights are required for the MS-VAR BCCI and MS-VAR GCCI. As we stated above, the aggregation procedure is implicit in the multivariate specification.

The latest release of both the MS-VAR GCCI and MS-VAR BCCI is shown in Figure 14.7, where they are compared to the reference dating chronology of the growth and classical business cycles, respectively.

All the seven slowdowns observed in the euro area since 1985 are correctly signaled by the MS-VAR GCCI. The average delay in detecting six out of the seven peaks of the growth cycle is 2.0 months. However, the MS-VAR GCCI falsely detects a slowdown between mid-2005 and mid-2006, which is not identified in the reference chronology. Moreover, the last slowdown stretches up to January 2015, that is almost two years after the last trough in the reference chronology (2013Q1). This constrasts with the GCCI whose last slowdown signal ended in February 2013.

The MS-VAR BCCI does not miss any of the three recessions suffered since 1985 nor it falsely signals any recession. The three peaks of the classical business cycle are detected with an average delay of 2.3 months, compared to the 7.0 months for the BCCI.

As summarized in Table 14.7, QPS and Concordance Index suggest that the MS-VAR BCCI is quite more accurate than the BCCI in locating the recessions in the euro area. However, the same statistics are pointing at a slightly greater accuracy of the GCCI compared to the MS-VAR GCCI in signaling slowdown of the growth cycle.

Finally, as pointed out above, turning points signals derived from the MS-VAR GCCI and MS-VAR BCCI are consistent with the ABCD approach.

Figure 14.7: MS-VAR BCCI (upper panel) and MS-VAR GCCI (lower panel) from August 1985 to April 2015, slowdowns of the growth cycles and recessions of the business cycle (both in grey shaded areas).
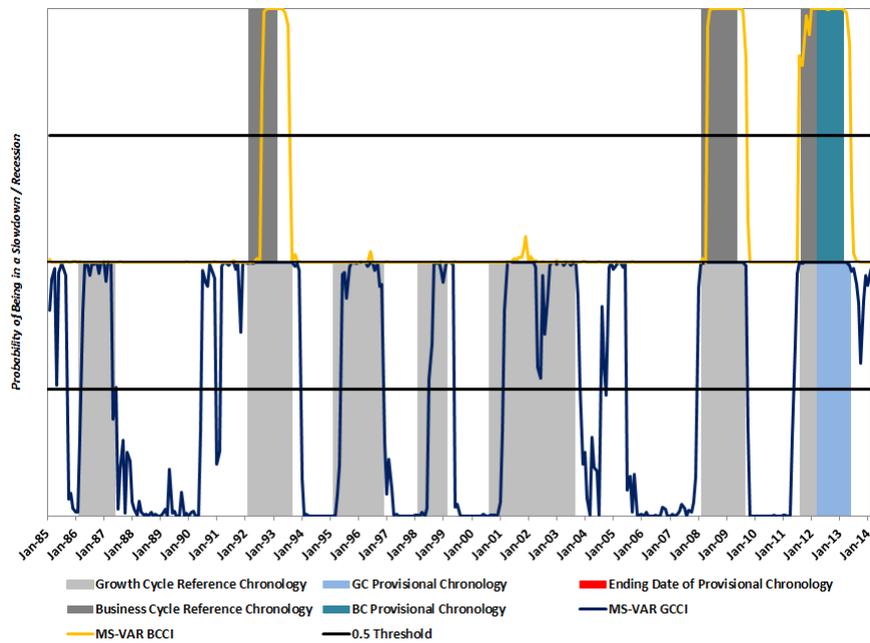
Table 14.7: MS-VAR GCCI's Accuracy Statistics

Indicator	QPS	Concordance Index
MS-VAR GCCI	0.22	0.74
MS-VAR BCCI	0.05	0.94

Member states extension

After the 2008-2009 global economic and financial crisis, we observed a decreasing tendency of the degree of synchronisation among euro area member countries, so that, by looking only at the euro area cyclical situation, the picture was less informative than before. For this reason, it was decided to extend the cyclical monitoring also at the euro area largest economies. Due to the positive results obtained with the multivariate models, we decided to adopt it also at country level while we decided not to produce an ACCI by country. The countries considered in a first phase were Germany, France, Italy, Spain, Belgium, the Netherlands and Portugal, while the construction of coincident indicators for the remaining euro area member countries is still ongoing. We adopt the same variable and model selection strategy as for the euro area but, in order to achieve better results we did some specific fine-tuning, keeping into account some countries' specificities. The model specification, as well as the selected variables, are shown in table 14.12.

It is worth noticing as in the model specification there are few exceptions with respect to the general specification adopted for the euro area which was used as the reference also at country level. The first one is the presence of five regimes in the models for Italy and Portugal, instead of the four usually adopted. This is probably due to the presence of stagnation phases for the Portuguese and Italian economies, which were not properly captured by the four regimes so that a specific one was requested. The second kind of exception is constituted by the fact that models for Portugal and Belgium do not include any heteroskedastic component which is determined by the fact that expansionary and recessionary phases for those countries were not as much asymmetric as for the other ones. Portugal is the country for which the model adopted is more distant from the other ones since it presents at the same time both the exceptions mentioned above.

Table 14.8: Model summary for the major MS

Country	Model	Recessions	Slowdown	Variables					
				IPI	UR	BUIL	IND	CONS	RETA
Belgium	MSI(4)-VAR(0)	R1	R1+R2	6	3	6	3	-	3
France	MSIH(4)-VAR(0)	R1	R1+R2	6	1	3	-	1	12
Germany	MSIH(4)-VAR(0)	R1	R1+R2	3	3	3	-	6	12
Italy	MSIH(5)-VAR(0)	R1	R1+R2	3	3	-	12	12	3
Netherlands	MSIH(4)-VAR(0)	R1	R1+R2	12	-	6	3	1	1
Portugal	MSI(5)-VAR(0)	R1+R2	R1+R2+R3	6	-	3	3	12	1
Spain	MSIH(4)-VAR(0)	R1	R1+R2	12	12	3	6	12	-

Starting from the indicators developed for the member states, we also derived an euro area pair of coincident indicators indirectly computed as a weighted average of the growth cycle and business cycle filtered probabilities returned by each country model where the weights are based on GDP share. Even if the indirect indicators are based on seven countries, since their GDP accounts for more than 80% of the euro area one, we can assume that the indicators actually computed are a very good proxy of the ones calculated based on all euro area countries. Figure 14.8 shows the behaviour of the indirect indicators for the growth cycle and the business cycle.

Figure 14.8: Indirect MS-VAR BCCI (upper panel) and Indirect MS-VAR GCCI (lower panel) from January 1992 to April 2015, slowdowns of the growth cycles and recessions of the business cycle (both in grey shaded areas)
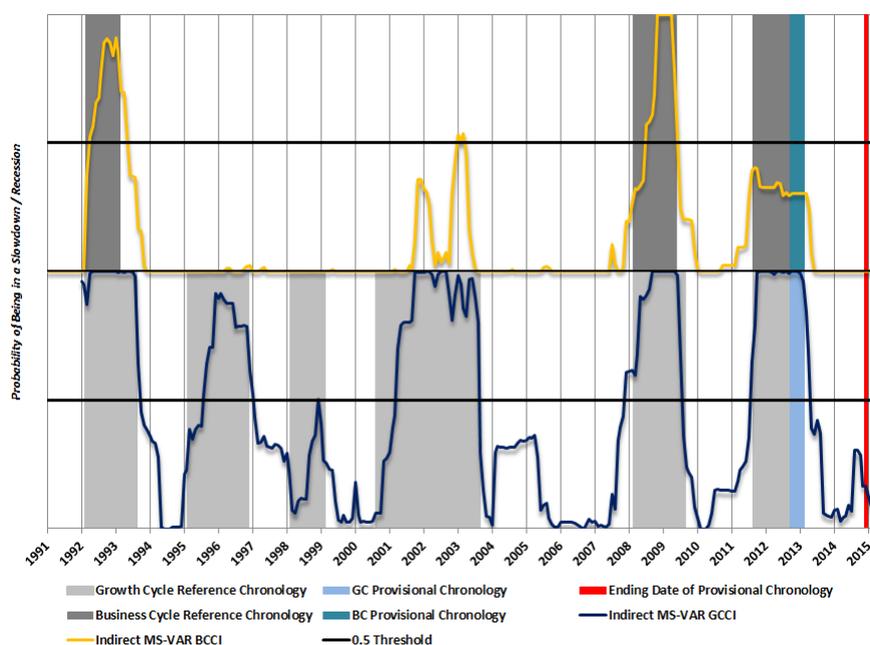
Based on the regular monthly production (complemented by an historical simulation exercise) of the euro area direct and indirect indicators as well as of the member countries ones, we have been able to identify several useful elements to evaluate the performance and the quality of them. In particular, we have concentrated our attention on the presence/absence of false signals and of missed cycles, as well as on the detection lag of peaks and troughs and on the degree of concordance between the indicators and the historical chronologies as measured by the Concordance Index and the Brier's Score. Tables 14.13 and 14.14 synthesize those elements for the MS-VAR GCCI and the MS-VAR BCCI respectively.

By looking at tables 14.13 and 14.14, we observe generally a good degree of concordance between the indi-

Table 14.9: Growth cycle outcome summary

Country	Slowdown missed	False slowdown	Average delay in locating slowdowns start (in months)	Accuracy in signalling slowdown	
				Brier's Score (QPS)	Concordance Index
Belgium	0	1 (2005)	0.7	0.22	0.76
France	0	0	2.8	0.18	0.82
Germany	1 (1998)	0	2.3	0.26	0.73
Italy	0	0	4.2	0.24	0.76
Netherlands	1 (1995-1996)	0	1.4	0.16	0.82
Portugal	0	3	0.8	0.18	0.80
Spain	1 (1997-1998)	0	3.5	0.24	0.76
EA direct	0	1 (2004-2005)	2.0	0.22	0.74
EA indirect	1 (1998)	0	3.0	0.10	0.87

Table 14.10: Business cycle outcome summary

Country	Recessions missed	False recessions	Average delay in locating peaks (in months)	Accuracy in signalling slowdown	
				Brier's Score (QPS)	Concordance Index
Belgium	1	0	7.3	0.13	0.87
France	0	0	2.5	0.02	0.98
Germany	0	1 (2001-2002)	2.7	0.08	0.92
Italy	1 (2001)	0	1.8	0.16	0.84
Netherlands	0	0	3.5	0.17	0.83
Portugal	0	0	4.3	0.12	0.88
Spain	0	0	2.3	0.05	0.94
EA direct	0	0	2.3	0.05	0.94
EA indirect	1 (2011-2013)	0	2.5	0.06	0.90

cators and the chronologies even if, especially for the euro area, some discrepancies emerged, especially in the latest period, let's say after 2011. At the euro area level both direct and indirect indicators perform similarly, at least until 2011. Indirect BCCI missed the last euro area recession while the direct one correctly detected it. For the GCCI, one missed cycle appears for the indirect indicator while one false signal appears for the direct one. When looking at the member countries indicators, we notice that both MS-VAR GCCI and MS-VAR BCCI for France are performing very well, with the only exception of a false signal recorded by the MS-VAR GCCI. Also for Germany and the Netherlands, the MS-VAR BCCI performs very well while there is room for improvement for the MS-VAR GCCI. Some improvements are required for the MS-VAR GCCI of Portugal to reduce the number of false signals and for Italy and Spain to increase timeliness. Concerning the MS-VAR BCCI, improvements are required for Belgium in order to reduce the number of missed cycles and, again for Belgium and Portugal, for increasing the timeliness.

The latest turning points based on the MS-VAR GCCI and MS-VAR BCCI for euro area and member countries as well as the ACCI, only ofr the euro area, are reported in Table 14.15.

Table 14.11: Growth, Business and Acceleration Cycle datings

Region	Method	Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough
Euro-Area	Growth cycle Provisional dating			2008Q1	2009Q3			2011Q3	2013Q2		
	Direct MS-VAR GCCI			12/2007	09/2009			05/2011			01/2015
	Indirect MS-VAR GCCI			11/2007	07/2009			07/2011	04/2013		
	Business cycle Provisional dating			2008Q1	2009Q2			2011Q3	2013Q1		
	Direct MS-VAR BCCI			04/2008	09/2009			07/2011	05/2013		
	Indirect MS-VAR BCCI			06/2008	06/2009						
	Acceleration cycle Provisional dating			2006Q2	2009Q1	20110Q2			2012Q4		
	Direct MS-VAR ACCI			06/2006	03/2009	12/2010	12/2011	03/2012	10/2012	12/2013	01/2015
Germany	Growth cycle			05/2008	07/2009			09/2011	03/2013		
	Business cycle			10/2008	06/2009						
France	Growth cycle			08/2007	07/2009			07/2011	08/2013		
	Business cycle			06/2008	05/2009						
Italy	Growth cycle			11/2007	06/2009			06/2011	05/2013		
	Business cycle			11/2007	04/2009			07/2011	04/2013		
Spain	Growth cycle			08/2007	11/2009			06/2010	08/2013		
	Business cycle			01/2008	11/2009			06/2011	04/2013		
Netherlands	Growth cycle	12/2005	07/2006	05/2008	12/2009			02/2011	02/2013		
	Business cycle			9/2008	12/2009			02/2011	10/2011		
Belgium	Growth cycle			08/2007	09//2009			04/2011	06/2012		
	Business cycle			07/2008	04/2009			05/2012			
Portugal	Growth cycle			06/2007	05/2009			06/2010	07/2013		
	Business cycle			05/2008	05/2009			09/2010	03/2012		

The table confirms the prevailing idiosyncratic behaviour of the euro area member countries, especially since 2011, accompanied by a low degree of synchronisation and diffusion of turning points. In particular, France and Germany did not enter in recession during the 2011-2013 period, while the Netherlands was only marginally affected by it. The remaining countries experienced a recession in 2011-2013 including the euro area (following the direct MS-VAR BCCI). All the euro area and its member countries experienced a slowdown phase started in 2011. Those elements show how the lack of synchronisation and diffusion is greater for the business cycle than for the growth cycle.

14.5 Conclusions

The construction and the maintenance of turning points composite indicators is a very complex task going through different steps. In this chapter we have describes in details such steps, providing either methodological or an empirical justification aiming to provide guidance for the compilation of turning points composite indicators. In particular, we have stressed how, in this specific case, it is not compulsory focusing on one reference cycle but, by referring to the ABCD sequence or to the $\alpha AB\beta CD$ one, it is possible to simultaneously monitoring alternative cycles. Furthermore, we have stressed the importance of regularly maintaining turning points chronologies in order to dispose of a benchmarking reference for the composite indicators which allows for a continuous real-time monitoring of their performance. In the chapter we have also shown as composite indicators can be compiled by using various specifications within the family of the Markov-Switching models.

The outcomes of the indicators is very sensitive to the model specification so that it is essential to carefully chose the most appropriate one in relation to the objectives and the needs that they have to achieve and serve. In this respect, despite its computaional complexity, the multivariate Markov-Switching models have proven to be a powerful instrument for constructing composite indicators which always respects the ABCD sequence. Nevertheless, under the condition that the coincident indicators for growth and business cycles, based on univariate modelling of component variables have the same degree of timeliness, this approach can become again very appealing.

14.6 Annex

14.6.1 Introduction

The compilation of cyclical composite indicators for the detection or forecasting of turning points, as described in this chapter 14, can be a quite complex process. During this process, several important decisions have to be taken which will affect substantially the overall quality of the process and the validity of its results. For this reason, in this annex I am proposing an operational scheme for the construction and, validation and maintenance of cyclical turning points indicators. Similar step by step approaches have already been presented in Mazzi and Montana (2009) and inMazzi et al. (2017), focusing respectively on the univariate BCCI and GCCI and on the multivariate MSVAR-BCCI and MSVAR-GCCI.

There are some innovative features which characterise the approach proposed in this annex with respect to the other already mentioned schemes. The first one is that here I am merging the schemes for the univariate indicators and for the multivariate ones into a single sequential process. The second is that here, together with a detailed description for each step, I am also proposing some suggestions and recommendations for the compilers. Finally, I have added a couple of steps providing guidance on how to ensure a regular monitoring of the indicators' performance and about the revision policy to be followed.

This annex is structured as follows: Section 14.6.2 describes the preliminary steps to be followed for the construction of turning points indicators; section 14.6.3 is devoted to the presentation of steps related to the modelling strategy and model selection aspects, while section 14.6.4 will describe the validation steps leading to the identification of the best indicator(s). Section 14.6.4 also proposes steps related to the implementation of a regular monitoring and revision strategy for the turning points indicators developed in previous steps, and discusses some dissemination issues. Section 14.6.5 will contain some concluding remarks.

14.6.2 Preliminary steps to the construction of turning points indicators

In this section we are presenting the first 3 steps for the construction of turning points indicators. In such steps, and especially in the first two, crucial decisions are required which will influence and conditioning the decisions to be taken in the subsequent steps.

Step 1 - Identification of the cycle to be monitored

(A) Description

Prior to any other action, business cycle analysts have to identify the cycle(s) to focus on. This decision is typically influenced by some factors stemming from the user's needs, policy-makers request, characteristics of the national economy, data availability and so on. Traditionally analysts have focused their attention on three main definitions of cycles:

1. Classical Business cycle (Burns and Mitchell definition; 1946), which is very relevant for detecting recessions but not very informative during usually quite long expansion phases.

2. Growth cycle (Output gap), which is very relevant to understand the position with respect to the potential output (trend) and more informative also during the expansion phases of business cycle. It leads to the peaks and troughs of the business cycle but it doesn't detect the start and the end of recessions. The output gap is also considered very relevant, especially by central banks, to anticipate inflation rate based on the modern Phillips curve relating inflation and output gap. The main weak point of the growth cycle is represented by the fact that the choice of alternative de-trending techniques can be substantially

affect the shape and the location of turning points. Furthermore, de-trending filters can produce very unstable growth cycle estimates at the end of the series.

3. Growth rate cycle (Acceleration cycle), which is characterised by the highest number of fluctuations and a high degree of volatility. It leads to the growth cycle peaks and business cycle troughs corresponding to the inflexion points of the classical business cycle. They determine the acceleration and deceleration phases of the economy.

In addition, it is also possible jointly monitoring more cycles as proposed by Anas and Ferrara (2004a,b). In this paper they propose to jointly follow the growth and business cycles within an integrated framework. This framework is defined by the 4 turning points associated to the two cycles and by the logical sequence of peaks and troughs. Such logical sequence of turning points assumes that peaks of the growth cycle anticipate those in business cycle and troughs in business cycle anticipate them in the growth cycle. Consequently, the framework has been defined by the sequence ABCD where A and D are respectively peak and trough of the growth cycle and B and C the peak and trough of business cycle.

In the same papers, the authors also include an extended version of this framework incorporating the acceleration cycle. In this case, the logical sequence of turning points becomes αABβCD where α and β are respectively the peak and the trough of the acceleration cycle.

Based on the Eurostat experience I consider that the monitoring of the cyclical economic situation is ensured by the simultaneous follow-up of the growth and the business cycles within the ABCD framework. The extension of the framework to include the acceleration cycle should provide more insight on the accelerating or decelerating behaviour of the economy, which can be very relevant especially in periods of slow or moderate growth. On the other hand, the risks associated to the high volatility of the acceleration cycle have to be carefully evaluated. Obviously, this approach can be quite resource consuming and in a first phase a country can decide to follow an easier strategy focusing on just a cycle.

(B) Recommendations

Key recommendations are the following:

1. Privilege as much as possible the adoption of the simultaneous monitoring of business and growth cycle within the ABCD framework with possible extensions to the $\alpha AB\beta CD$ one.

2. In case where resource constraints or problems in data availability do not allow for implement the recommendation 1, a single indicator should be developed instead. In this case, also based on the Eurostat experience, the suggestion is to concentrate on the growth cycle since it is more informative especially for developing and emerging economies.

Step 2 - Constructing a historical dating chronology

(A) Description

A historical dating chronology for each of the cycle selected in the previous step is an essential tool to benchmark the turning points indicators to be developed in the following steps. Without a historical chronology, the validation of turning points indicators could not be performed from a statistical point of view lowering considerably their relevance and compromising the methodological soundness of the whole constructing process.

For the reasons mentioned above, it is crucial at this stage to search for the existence of official historical dating chronologies, if any, or for other reference chronologies regularly updated and maintained, based on sound methodologies. If such chronologies do not exist or do not cover, totally or partially, the cycle's definitions to be monitored, before computing turning points indicators, they have to be constructed possibly adopting a simple dating rule. Statistical dating chronologies should cover a time horizon as long as possible, also depending

on data availability. They should be constructed with the aim of keeping past turning points, after a certain number of years, fixed, even we have to be aware of the difficulty of such objective, especially in presence of big revision in official statistics.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. Use the quarterly GDP plus one or two more monthly key variables such as industrial production, and unemployment rate as reference variables from the historical dating chronologies.

2. Use a simple non-parametric dating rule such as the Hardin and Pagan (2002) to identify past turning points;

3. Even if in step 1 it has been decided to focus on just one cycle, compute the historical dating chronologies for all 3 cyclical definitions. This will be not very complex and the benefits very relevant for future analysis;

4. When constructing the chronologies for various cycles, ensure that the ABCD sequence or its extended version $\alpha AB\beta CD$ are fulfilled.

Step 3 - Construction of a dataset to be used for the construction of cyclical turning points indicators

(A) Description

A middle-sized dataset has to be developed at this stage. It will constitute the main statistical base for the following steps where data and model selection tasks will be performed. To allow the jointly monitoring of the business growth and acceleration cycles, the dataset should mainly contain the key macro-economic indicators, available possibly at monthly frequency, as well as opinion surveys data and, eventually, other soft and hard indicators which could be useful for detecting and forecasting turning points. If only growth and acceleration cycles will be targeted, then the data set should mainly contain soft data and only a few numbers of key macro-economic indicators. By contrast, if the business cycle is targeted, then there is no need to have opinion surveys data in the dataset which should contain only macro-economic indicators and, if needed, other relevant soft data such as financial variables.

The data set should contain the original values of the indicators as well as their most appropriate data transformation to highlighting cyclical movements. The dataset should contain in priority seasonally adjusted data but also non seasonal adjusted ones should be included mainly for advance research projects. Ideally the dataset should contain the vintages of the selected indicators recorded over a sufficiently long time horizon covering possibly two cycles at least.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. The dataset containing the selected indicators should be regularly updated and maintained, subject to a continuous quality monitoring and be regularly backed up in order to create historical vintages if not previously available.

2. The size, composition and structure of the dataset should target the cycles to be monitored as in step 1.

14.6.3 Steps towards the construction of turning points indicators

In this section, I am presenting the various aspects of the modelling strategies to be followed for constructing univariate cyclical turning point indicators as well as multivariate ones.

Step 4 - Variable selection

(A) Description

Starting from the middle-sized dataset developed in step 3, a variable selection process is necessary at this stage in order to identify small number of variables to be considered as candidate component of the cyclical turning points indicators to be constructed. Such selection will be mainly based on the ability shown by each variable in the dataset of timely and precisely detecting turning points within a simulation exercise against the non-parametric historical dating chronologies constructed in step 2. The best possible scenario for this exercise is represented by a real-time simulation by using as much as possible past vintages of the variables to be tested.

The ability of the variables to timely and precisely detect turning pints will be assessed by using simple statistical tools such as graphical investigation or the use of very simple time-series models to be fitted to each variable. Obviously, the specific characteristics of each series should be considered when deciding which kind of turning points they are supposed to be able to detect. For example, as already mentioned in the previous step, it will be inappropriate and even misleading to test the ability of detecting business cycle turning points by opinion surveys data which, by definition, measure only growth cycle and likely acceleration cycle.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. Since large scale based indicators do no necessarily perform better than small scale ones, and also considering that they can be more complex to maintain and monitor, the main suggestion is to identify a relatively small number of candidate variables to be used in the following steps.

2. Use all statistical knowledge about the series in the dataset, such as their characteristics, peculiarities etc., as priors in the variable selection exercise.

Step 5 - The modelling strategy

(A) Description

The modelling strategy for the construction of cyclical turning points indicators has to be modelled on the choices made in step1. If in step1 it has been decided to focus on just one cycle, either the business or the growth one (the acceleration could be much more rare as a choice), the chosen variables should be modelled by means of univariate model and the results should be aggregated by using a simple aggregation scheme.

In the case where the growth and the business cycle have been chosen to be monitored in step 1 then it is possible to evaluate competing alternatives. The first one is similar to the one proposed for constructing an indicator for a single cycle and it will consist of estimating independent indicators for each cycle by using univariate models fitted to each component and a simple aggregation scheme at the end. The alternative is a jointly multivariate modelling of the growth and business cycle which will avoid the aggregation step and the identification of the weighting scheme. If also the acceleration cycle was included step 1, for this cycle it is

necessary to proceed always with an independent modelling since, for purely mathematical reasons, it is not possible to simultaneously model growth business and growth cycle. So in this case there will be multivariate model for growth and business cycle and a univariate model for acceleration cycle.

Finally, it has to be noted that several non-linear techniques, as discussed in chapter 13, could be used for constructing turning points indicators. Nevertheless, based on the interaction evidence, and also on personal experience (see Billio, Ferrara et al 2013) I assume in the following of this annex, consistently to that done in chapter 14, that the Markov switching models are adopted by default.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. Provide a clear statement on the chosen modelling strategy and on its justifications.

2. Within the class of best performing models, privilege simple, easily replicable and economically interpretable ones.

Step 6 - Model specification: univariate case

(A) Description

In order to simplify the presentation, I am assuming here that all 3 cycles should be monitored by means of turning points indicators independently constructed based on univariate modelling of the component series. Obviously, this description also applies to the cases where only two or one cycle has to be monitored. For each cycle to be monitored the candidate series will be modelled by means of a univariate MS model:

$$MSIH(k) = AR(l)$$

Where H indicates the presence of the heteroscedastic components, K the number of regimes and L the number of lags of the autoregressive part. Each model should target one of the selected cycles: acceleration, business or growth cycle following the characteristics of the variable involved. Consequently, the number of regimes and the presence of heteroscedasticity should vary accordingly.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. Privilege univariate models with a number of regimes not exceeding 3, keeping in mind that, especially for the growth and the acceleration cycles, two regimes should usually be enough.

2. Models without autoregressive structure ($l = 0$) should be privileged.

Step 7 - Model selection: univariate case

(A) Description

From step 6, N univariate best fitting models are identified for each reference cycle. Each of them will return a probability of being in recession, slowdown or deceleration phase according to the reference cycle. For each reference cycle a number of turning points indicators will be then derived as a weighted average of the recessions/slowdown/deceleration probability returned by K components series where $k < n$.

In practical terms the outcome of this step will constitute of $M1$ indicators for the acceleration cycle, labelled ACCI ($M1$), $M2$ indicators for the growth cycle labelled GCCI ($m2$) and $M3$ indicators for the business cycle labelled as BCCI ($M3$). Formally this step will lead to the compilation of the following turning points indicators:

Main recommendations for this step can be synthesised as follows:

1. $M1$ acceleration cycle composite indicators ACCI of the form:

$$ACCI(j_1, t) = \sum_{i=1}^{k_1} h(i)p^{ac(i,t)} \qquad j_1 = 1 \ldots m_1, t = 1 \ldots T, \qquad (14.1)$$

K_1 is the number of variables included in the j_1 indicators, $h(i)$ are standardised weights so that $\sum_{i=1}^{k_1} h(i) = 1$ and $p^{ac(i,t)}$ are the deceleration probabilities returned by each component variable at time t. If $ACCI(j_1, t) > 0.5$ the economy is in a deceleration phase.

2. $M2$ growth cycle composite indicators GCCI of the form:

$$GCCI(j_2, t) = \sum_{i=1}^{k_2} w(i)p^{gc(i,t)} \qquad j_2 = 1 \ldots m_2, t = 1 \ldots T, \qquad (14.2)$$

$W(i)$ are standardised weights withe same properties as $h(i)$, k_2 is the number of variables included in the j_2 indicator and $p^{gc(i,t)}$ is the slowdown probability returned by each k_2 component series at time T. When $GCCI(j_2, t) > 0.5$ the economy is in a slowdown phase meaning that it is growing below the trend.

3. $M3$ business cycle composite indicators BCCI of the form:

$$BCCI(j_3, t) = \sum_{i=1}^{k_3} v(i)p^{bc(i,t)} \qquad j_3 = 1 \ldots m_3, t = 1 \ldots T, \qquad (14.3)$$

$v(i)$ is a set of standardised weights with the same properties of $h(i)$ and $W(i)$, k_3 is the number of variables included in the j_3 indicator and $p^{bc(i,t)}$ is the recession probability returned by each k_3 components at time t. If $BCCI(j_3, t) > 0.5$ the economy is in a recessionary phase.

In all cases 0.5 is usually the adopted threshold associated to the so called natural rule of Hamilton (1989). Since the growth cycle and the business cycle indicators, GCCI and BCCI, are independently compiled, there is no guaranty that when $BCCI > 0.5$, implies $GCCI > 0.5$, consequently the ABCD sequence is not automatically fulfilled.

(B) Recommendations

Main recommendations for this step can be synthesised as follows:

1. Since the 0.5 threshold has demonstrated in several empirical studies to be a good compromise between the flexibility and timeliness in detecting turning points on one hand and the risk aversion which could characterise such indicators, it should be adopted for all turning points indicators.

2. Keep the number of components for each indicator as much as possible limited, possibly not exceeding 5. This limitation will contribute to avoid that the procedures for constructing turning points become too heavy and difficult to be managed, maintained and interpreted.

3. Check the compliance of the BCCI(s) and the GCCI(s) indicators with the ABCD sequence and privilege those indicators which fulfil the sequence.

4. In case of inconsistency with the ABCD sequence, consider the possibility to move to a multivariate strategy as described in steps 8 and 9.

Step 8 - Model specification: multivariate case

(A) Description

Selected variables in step 4 are here used to identify and estimate a number of vector autoregressive Markov-Switching models (MS-VAR):

$$MSIH(K) - VAR(L), \tag{14.4}$$

where H indicates the presence of heteroskedasticity, (K) is the number of regimes and (L) the number of lags of the autoregressive part. This multivariate model aims to simultaneously produce turning points indicators for the business and growth cycle respectively. Dealing simultaneously with growth cycle and business cycle implies a number of regimes not smaller than 3 so $k => 3$, while the heteroskedastic part H can or cannot be present depending on the degree of asymmetry of fluctuations. Based on the Eurostat experience, the most commonly found number of regimes for these indicators is 4. Concerning the heteroskedasticity, its presence has been significantly detected in most multivariate models.

(B) Recommendations

1. Keep the number of regimes as much as possible small and easy to be justified and interpreted. The number of regimes exceeding 5 should be avoided either for the related computational complexity or for the difficult interpretation in economic terms.

2. Privilege models with $L = 0$ meaning not ar structure, unless there is an empirical significant evidence indicating that models with $l > 0$ outperform these with $l = 0$.

3. In case where $l > 0$, the number of lags should be kept as small as possible and not exceeding 3.

Step 9 - Model selection: multivariate case

(A) Description

From step 8, N best fitting models are identified, each of them producing a pair of coincident indicators for the growth cycle and the business cycle respectively, labelled as MS-VAR GCCI (multivariate growth cycle coincident indicator) and MS-VAR BCCI (multivariate business cycle coincident indicator):

$$MS - VAR\,GCCI(j) \text{ and } MS - VAR\,BCCI(j); \ j = 1 \dots n. \tag{14.5}$$

Each composite indicator is defined between 0 and 1, and can be viewed as a composite probability of being in a recessionary phase for the MS-VAR BCCI (j) and in a slowdown phase for the MS-VAR GCCI (j). The recession/slowdown regions are defined on the basis of a threshold, usually set up to 0.5. Obviously higher values for the threshold as well as smaller ones can be used. In the first case, the indicators should detect less fluctuation, missing some cycles, while in the second one they will detect more fluctuations than really occurring cycles (see chapter 14 for more details). By adopting the 0.5 threshold we can have then the following cases:

- MS-VAR BCCI (j) > 0.5 = recession

- MS-VAR GCCI (j) > 0.5 = slowdown

- MS-VAR BCCI (j) < 0.5 =recovery

- MS-VAR GCCI (j) < 0.5 = expansion

One of the most interesting features of this multivariate class of models is that by construction,

$$MS - VAR\,BCCI(j) > 0.5 \text{ and } MS - VAR\,GCCI(j) > 0.5, \tag{14.6}$$

so that the ABCD sequence is always fulfilled.

(B) Recommendations

1. Since the 0.5 threshold has demonstrated in several empirical studies to be a good compromise between the flexibility and timeliness in detecting turning points on one hand and the risk aversion which could characterise such indicators, it should be adopted for all turning points indicators.

2. Keep the number of components series as small as possible and possibly not exceeding 5. In addition to the advantages mentioned in recommendation 2 of step 8, in the multivariate case, this will contribute to reduce the risks of lack of convergences of the estimation algorithms.

14.6.4 Model validation and the identification of the best indicator(s)

In this section I am describing the validation process of the indicators developed in step 7 (univariate indicators) and in step 9 (multivariate indicators). Later, the main criteria allowing for the identification of the most performing indicators for each group will be presented.

Step 10 - Dynamic comparison of the turning points indicators

(A) Description

Within a dynamic simulation exercise, the $M(i)\,i = 1, 2, 3$ univariate indicators developed in section 7 for the acceleration, growth and business cycle, as well as the N pair of multivariate composite indicators for the growth and the business cycle are compared with the non-parametric historical turning point dating developed in step 2. The time span for this comparative exercise should be long enough and contain a certain number of cycles in order to allow for the identification of strong and weak points of each indicator in detecting turning points. Obviously, the best possible way to conduct this exercise is in real-time, by using historical vintages instead of final vintages.

(B) Recommendations

1. Select a time span possibly not shorter than 15-20 years including at least 2-3 recessionary events and a higher number of other cyclical movements.

2. Privilege the use of historical vintages whenever available even for a subset of the selected time span.

Step 11 - Identification of the best performing indicators

(A) Description

The identification of the best performing ACCI, BCCI and GCCI as developed in step 7, as well as of the best pair of MS VAR-BCCI and MS VAR-GCCI, as developed in step 9, is based on the outcome of step 10, using the following statistical criteria:

- Average lead/lag in identifying peaks;

- Average lead/lag in identifying troughs;

- number of false cycles detected;

- number of missing cycles;

- Concordance Index;

- Brier's Score (QPS).

In an ideal situation, in case of coincident indicators the average lag in detecting turning points should be 0 or very small, while in the case of leading indicators the average lead in anticipating turning points should be positive. Furthermore, we should also expect that both the number of false signals as well as the number of missing cycles should be 0. Nevertheless, such an ideal situation is not achievable since the number of false signals and the number of missing cycles have to be viewed as the type2 and type1 errors of the estimation process. In statistical inference, we have learned that it is impossible to minimise simultaneously both errors and that a choice had to be made.

Finally, still in an ideal case a concordance index equal to 100 or very close and a QPS near 0 can be expected.

(B) Recommendations

1. Since the risk associated to the announcement of a non-existing cycle is considered higher than the one associated to missing a cycle, I recommend privileging the indicators minimising type 2 error.

2. Indicators characterised by high delays in the identification of turning points, especially peaks, should be discarded because they will not be useful in practice.

3. Indicators showing a constant behaviour over the simulation period should be preferred to these with a highly variable performance.

(5) - Regular monitoring of the turning points indicators

In this section we provide guidance on the regular assessment of the indicators' performance as well as on the strategy to be adopted to revise indicators on regular bases.

Step 12 - Regular assessment of the indicators' performance

(A) Description

The same criteria used to select the best performing indicator(s) in the previous step, can be used here to regularly assess the performance of the selected indicator(s), in particular to discover if it is subject to any kind of deterioration during the time. In chapter 14, summary tables have been proposed to compare the behaviour of indicators across countries. In these tables, one for each cycle, together with the indicator model specification, all 6 statistical criteria presented in step 11 are displayed. This is a very useful and powerful tool to evaluate on a regular base the performance of the chosen indicator(s) and to highlight possible risks of deterioration.

(B) Recommendations

1. Implement for each cycle (acceleration, growth and business cycle) and for each indicator a table following the scheme presented in chapter14 and regularly update it.

2. Update the table at least twice a year but ideally every quarter.

Step 13 - Revisions

(A) Description

When constructing composite indicators for detecting or anticipating turning points, it is essential to ensure a stability of the signals over the time. This implies a high stability of subsequent vintages of the indicators. In practical terms, there are several factors which can affect such stability. The first one is related to the regular revision process which characterises most macroeconomic variables. Unfortunately, this process is not on the hands of the compilers of turning points indicators but of data producers. Nevertheless, routine revision process is quite smooth and rarely produces changes able to affect the signals returned by composite indicators. Such uncommon events are generally associated to the occurrence of big shocks such as at the beginning of the global financial and economic crisis where data have been subject to huge revisions.

Other elements that can impact the stability of the signals returned by composite turning points indicators is constituted by the decisions taken by the compilers concerning the re-estimations and re-specification strategies for the indicators. Re-estimating and re-specifying too often the models is obviously improving the precision of the latest estimates but the counterpart is a high degree of instability over the times which can confuse policy-makers and analysts. On the other hand, never re-estimating and re-specifying the models, is progressively lowering their ability to timely and reliably detect turning points. An intermediate solution able to find a balance in the trade-off between timeliness and reliability from one hand and the stability on the other hand is necessary. The Eurostat experience demonstrated that a conservative approach privileging the stability of the indicators over the time is preferable.

(B) Recommendations

1. Establish a regular contact with data producers in order to better understand the regular revision process characterising the macroeconomic variables included in composite indicators for turning point detection.

2. Re-estimate the model parameters on yearly basis unless relevant anomalies in the data (e.g. very significant outliers, level shift, unexpected revisions, etc.) occur during the year. In such case, the need for an exceptional re-estimation of the models has to be evaluated.

3. Re-specify and re-identify the models usually every 5 years or whenever major revisions (e.g. changes in methodology definition and classification of macroeconomic indicators) occur. In such cases, the need for the re-specification and the re-identification of the models has to be evaluated.

(6) Dissemination of composite indicators for turning point detection

Step 14 - Dissemination issues

(A) Description

The most logical way to disseminate turning point indicators is the traditional one based on the presentation of their behaviour using tables and graphs to which regular press releases could be also associated. Neverthe-less, the interpretation of the turning points indicators is not so evident especially for non-expert users. This interpretation can become even more complex when multiple indicators are released for example for the simul-taneous detection of turning points indicators for the growth and business cycle within the ABCD approach. In order to simplify the reading of the indicators and to provide a synthetic message it is then advisable to complement the usual dissemination way with graphical or visual dissemination tool explicitly designed for

business cycle analysis. All turning points indicators have also to be properly documented within a standard metadata framework agreed at the international level.

(B) Recommendations

1. Complement the standard dissemination of composite indicators for turning point detection with advanced graphical and visualisation tools able to simplify their interpretation and to give a clear picture of the business cycle situation.

2. Provide a standard metadata file for each turning point indicators in line with the international standards and complement it, if needed, by more technical documentations and papers.

14.6.5 Conclusions

In this annex, I have generalised and completed the step-by-step approaches already presented in Mazzi and Montana (2009) and Mazzi et al. (2017). The step-by-step approach proposed here tries to cover all phases in the construction of composite indicators for turning point detection, starting from the decision on what to compile to the compilation process until the dissemination of the indicators. The recommendations provided at each step are generally based on methodological considerations also supported by the experience that Eurostat has made in the compilation of such indicators since 2007.

Bibliography

Anas J., Ferrara, L. (2004a), Turning points detection: The ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, 1, 1-36.

Anas J., Ferrara L. (2004b), A comparative assessment of parametric and non parametric turning points detection methods : the case of the Euro-zone economy, in Monographs of Official Statistics : Statistical Methods and Business Cycle Analysis of the Euro zone, G.L. Mazzi and G. Savio (eds.), Eurostat, pp. 86-121.

Anas J., Billio M., Ferrara L., LoDuca M. (2007a), A turning point chronology for the Euro-zone, in Growth and Cycle in the Euro zone, G.L. Mazzi and G. Savio (eds.), 261-274, Palgrave MacMillan, New York.

Anas J., Billio M., Ferrara L., LoDuca M. (2007b), Business cycle analysis with multivariate Markov Switching models, in Growth and Cycle in the Euro zone, G.L. Mazzi and G. Savio (eds.), 249-260, Palgrave MacMillan, New York.

Anas J., Billio M., Ferrara L., Mazzi G.L. (2008), A system for dating and detecting cycles in the Euro area, The Manchester School, 76, 549-577.

Artis M., Marcellino M., Proietti T. (2004), Dating business cycles: A methodological contribution with an application to the Euro area, Oxford Bulletin of Economics and Statistics, 66, 537-565.

Artis M., Krolzig H-M., Toro J. (2004), The European business cycle, Oxford Economic Papers, 56, 1-44.

Banerji A. (1999), The three Ps: Simple tools for monitoring economic cycles, Business Economics, 34, 72-76.

Bellégo C., Ferrara L. (2012), Macro-financial linkages and business cycles: A factor-augmented probit approach, Economic Modelling, 29, 1793-1797.

Bengoechea P., Camacho M., Perez-Quiros G. (2006), A useful tool to identify recessions in the Euro-area, International Journal of Forecasting, 22, 735-749.

Brier G.W. (1950), Verification of Forecasts Expressed in Terms of Probability, Monthly Weather Review, 78, 1-3.

Bruno G., Otranto E. (2008), Models to date the business cycle: The Italian case, Economic Modelling, 25, 899-911.

Bry G., Boschan C. (1971), Cyclical Analysis of Time series: Selected Procedures and Computer Programs, NBER, Columbia University Press.

Burns A.F., Mitchell W.C. (1946), Measuring Business Cycles, NBER, Columbia University Press.

Camacho M., Perez-Quiros G. (2010), Introducing the euro-sting: Short-term indicator of euro area growth, Journal of Applied Econometrics, John Wiley & Sons, Ltd., vol. 25(4), pages 663-694.

Camacho M., Perez-Quiros G., Saiz L. (2008), Do European business cycles look like one?, Journal of Economic Dynamics and Control, 32, 2165-2190.

CEPR (2003), Euro Area Business Dating Committee, Press Release, September 22, 2003.

CEPR (2009), Euro Area Business Dating Committee, Press Release, March 31, 2009.

Chauvet M. (1998), An econometric characterization of business cycle dynamics with factor structure and regime switching, International Economic Review, 39, 969-996.

Chauvet M., Hamilton J. (2006), Dating business cycle turning points in real time, in Nonlinear Time Series Analysis of Business Cycles, Milas C., Rothman P., Van Dijk D. (eds.), Elsevier Science, Amsterdam.

Darné O., Ferrara L. (2009), Identification of slowdowns and accelerations for the euro area economy, CEPR Discussion Papers 7376, C.E.P.R. Discussion Papers.

Darné O., Ferrara L. (2011), Identification of slowdowns and accelerations for the euro area economy, Oxford Bulletin of Economics and Statistics, 73, 3, 335-364.

Demers F., MacDonald R. (2007), The Canadian business cycle: A comparison of models, Working Paper No. 2007-38, Bank of Canada.

Diebold F.X., Rudebusch G.D. (1989), Scoring the leading indicators, The Journal of Business, 62, 3, pp. 369-391.

Estrella A., Mishkin F.S. (1998), Predicting U.S. Recessions: Financial Variables As Leading Indicators, The Review of Economics and Statistics, MIT Press, vol. 80(1), pages 45-61, February.

Estrella A., Rodrigues A.P., Schich S. (2003), How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States, The Review of Economics and Statistics, MIT Press, vol. 85(3), pages 629-644, August.

Ferrara L. (2003), A three-regime real-time indicator for the US economy, Economics Letters, 81, 373-378.

Ferrara L., Mazzi G.L. (2016), Parametric models for cyclical turning points indicators, *this volume*.

Forni M., Hallin M., Reichlin L. (2005), The generalized dynamic factor model: one-sided estimation and forecasting, Journal of the American Statistical Association, 100, 830-840.

Hamilton J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57, 357-384.

Hamilton J.D. (1990), Analysis of time series subject to changes in regime, Journal of Econometrics, 45, 39-70.

Hamilton J. (2003), What is an oil shock?, Journal of Econometrics, 113, 363-398.

Harding D. (2004), Non-parametric turning point detection, dating rules and the construction of the Euro-zone chronology, in Monographs of Official Statistics: Statistical Methods and Business Cycle Analysis of the Euro zone, G.L. Mazzi and G. Savio (eds.), Eurostat, 122-146.

Harding D., Pagan A. (2002), Dissecting the cycle: A methodological investigation, Journal of Monetary Economics, 49, 365-381.

Harding D., Pagan A. (2006), Synchronization of cycles, Journal of Econometrics, 132, 59-79.

Kauppi H., Saikkonen P. (2008). Predicting U.S. Recessions with Dynamic Binary Response Models, The Review of Economics and Statistics, MIT Press, vol. 90(4), pages 777-791, November.

Kim C-J., Nelson C.R. (1998), Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, Review of Economics and Statistics, 80, 188-201.

Krolzig H.M. (1997), Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Applications to Business Cycle Analysis, Springer, Berlin.

Mazzi G.L., Savio G. (2007), Growth and Cycle in the Eurozone, Palgrave-MacMillan, New York.

Mazzi G.L., Mitchell J. and Ozyildirim A (2017), Chapter 9: Alternative de-trending methods, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mazzi G.L., Moauro F. and Ruggeri-Cannata R. (2011). A statistical overview of the economic situation in the euro area, 46th Scientific Meeting of the Italian Statistical Society.

Mazzi G.L and Montana G. (2009). A system of rapid estimates to improve real-time monitoring of the economic situation: the case of the Euro Area, National Institute Economic Review n.210, October 2009, 63-70.

Mintz I. (1969), Dating postwar business cycles: Methods and their application to Western Germany, 1950-67, Occasional Paper No. 107, NBER.

Mönch E., Uhlig H. (2005), Towards a monthly business cycle chronology for the Euro area, SFB Discussion Paper No 649, April 2005.

Montoya, L. and de Haan, J. (2008), "Regional business cycle synchronization in Europe?," International Economics and Economic Policy, Springer, vol. 5(1), pages 123-137, July, available online at: https://ideas.repec.org/a/kap/iecepo/v5y2008i1p123-137.html.

Newey W., West K. (1984). A simple, positive semi-definite, heteroskedastic and autocorrelated consistent covariance matrix.

Rudebusch G.D., Williams J.C. (2009), Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve, Journal of Business & Economic Statistics, American Statistical Association, vol. 27(4), pages 492-503.

Schirwitz B. (2009), A comprehensive German business cycle chronology, Empirical Economics, 37, 2, 287-301.

Stock J., Watson M. (2002), Has the Business Cycle Changed and Why?, NBER Working Papers 9127, National Bureau of Economic Research, Inc.

Stock J., Watson M. (2003), How did leading indicator forecasts perform during the 2001 recession?, Economic Quarterly, Federal Reserve Bank of Richmond, issue Sum, pages 71-90.

Zarnowitz V., Ozyildirim A. (2006), Time series decomposition and measurement of business cycles, trends and growth cycles, Journal of Monetary Economics, 53, 1717-1739.

V Growth Composite Indicators

15

An Overview of Growth Composite Indicators

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ludovic Calès
European Commission, Joint Research Centre (JRC)
ludovic.cales@ec.europa.eu

Table of Contents

15.1 Introduction

The composite indicators presented in this part of the handbook are substantially different from those discussed in Parts 3 and 4. Indeed, they are not focusing on estimating in real-time or anticipating cyclical movements but rather the short term evolution of a given key macroeconomic indicators such as GDP, inflation, etc. The short-term evolution of some key macroeconomic indicators can be defined in several ways according to the characteristic of the series, to the use of seasonally or not seasonally adjusted data, etc. By consequence, indicators should be built up taking into account which is the key measure of evolution usually considered for the variable of interest. In terms of their interpretation, those composite indicators can be seen, as an alternative to traditional nowcasting and forecasting techniques, to either fill gaps in timeliness or provide one or two-step-ahead forecasts. The main difference with nowcasting and forecasting techniques is that, in this specific case, the statistical variables with now/forecasting power are first synthesised in a composite indicator which is then used in the purely forecasting exercise. It is useful to notice that there are several coincident or leading indicators which are regularly produced and disseminated by several public and private institutions, mainly aiming to forecast GDP trends, but also prices, etc.

The chapter is structured as follows: Section 15.2 describes the selection of the reference variable, section 15.3 elaborates on how short-term pattern of a given variable can be measured. Section 15.4 outlines the methdology commonly used for nowcasting and forecasting. Section 15.5 reviews the relevant literature in this field. Section 15.6 provides an overview of some regularly available coincident and leading indicators. Section 15.7 provides a guide to the part. Section 15.8 concludes.

15.2 The selection of the reference variable

For this category of indicators, the selection of the target variable is mainly determined by the need of providing a picture of the economic situation for the real-time and/or for the near future. By consequence, the GDP is the most commonly used reference variable even if there are also examples where price, employment, industrial production and others are considered. When we are considering the identification of the reference variable, we cannot only consider timeliness aspects but also other relevant ones such as the volatility and the overall quality of the variables. In particular, a high degree of volatility of a given reference variable could make quite complex the construction of a good composite indicator. As an example, see Charpin and Mazzi (2017), despite several attempts to construct a composite indicator to anticipate the industrial production for the euro area, the results have been considered quite disappointing due to the high degree of volatility of the euro area industrial production. Furthermore, the presence of outliers, structural breaks and other irregularities can also make quite complex the construction of a good quality composite indicator. Finally, when selecting a target variable we have also to consider the availability of the reliable candidate variable to be used in the construction of a composite indicator.

The reference variable in this section can also be represented by a latent one which is strongly related to an observed relevant macroeconomic variable. This is the case of a monthly target variable, representing either the monthly evolution of GDP or employment, while those variables are only available at quarterly frequency. In this particular case, the aim of the composite indicators is more related to the increase of the time coverage than to the increase of the timeliness. An example of composite indicators aiming to estimate a monthly proxy of GDP is provided in Frale et al. (2010) and Frale et al. (2011).

15.3 The definition of the short-term pattern

As mentioned in Section 15.1, the short-term pattern of a given variable can be measured in several ways. In principle, composite indicators presented in this section should try to estimate in real-time or to anticipate the most commonly used measures of the short-term evolution for each variable. For example, for seasonally adjusted flow data, the short-term evolution is usually measured by the period-on-period growth rate, while for non-seasonally adjusted ones, by the year-on-year growth rate. By consequence, composite indicators of GDP and industrial production index, usually mimic the present and near future evolution of their period-on-period growth rate, while the composite indicator of inflation should mimic the annual growth rate of consumer prices. Nevertheless, the presence of high degree of volatility, especially for industrial production index can, sometimes, suggest a further smoothing by applying a short moving average before or after the computation of growth rate.

For stock variables, such as unemployment or employment, both rates (expressed in terms of another variable such as the active population) or period-on-period growth rates are targeted. Obviously, this list is not at all exhaustive and there are other ways to express the reference variables. For example, in the US annualised growth rates are often used instead of period-on-period ones. In order to provide a consistent message to policy makers and analysts, it is important that the composite indicators developed in this section mimic the pattern usually chosen by official statisticians to communicate the results of their key macroeconomic variables. This decision does not influence only the communication of the real-time estimates or the future expectations derived by the use of composite indicators but also the ways in which the models used in this part are specified. It is by the way quite well known that the best model specification achieved in level terms does not necessarily produce the best possible estimates of the growth rate derived by such levels.

15.4 Methodology

The most commonly used methodologies for nowcasting and forecasting, which constitute the basic blocks for the construction of composite indicators, are presented in this part. They cover a large variety of methods dealing with the dimensionality problem of the dataset and the representation of the economy, either separately or simultaneously.

The methods dealing with the two problems separately focus first on the variables entering the dataset which are either selected or reduced. The variable selection methods intend to select the best subset of predictors from a large dataset in order to avoid collinearity and the noise induced by unnecessary predictors. These methods include the backward elimination, consisting in removing the predictors according to their significance and stepwise regressions, allowing for the re-insertion of predictors. They can also be based on information criterion such as the Akaike and Bayes information criterion. An extension of the stepwise regression, the LARS algorithm, is presented in Chapter 18. The variable reduction methods intend to summerize the information of a large dataset into few synthetic variables, usually independent. These methods include the principal components analysis, the partial least squares and the Bayesian shrinkage regression. These methods are presented in Chapter 5. The second step focuses on the modelling. They are usually a combination of the following methods:

- the usual linear models such as VAR, VECM, etc.

- the bridge models provide equations relating low-frequency variables and time aggregated high-frequency variables. The forecasts of the high frequency variables are obtained by specific high frequency time series models. They are then aggregated and inserted in the bridge equations to obtain the forecast of the low frequency variable. Chapter 18 considers a bridge error correction model.

- the mixed data sampling methods (MIDAS and U-MIDAS) consist in very parsimonious distributed lag models where the low-frequency is regressed upon higher-frequency explanatory variables

The methods dealing with the dimensionality problem and the modelling simultaneously are usually casted in the state-space representation. They include:

- the dynamic factor models which assume that few latent factors drive a large number of the economic variables. Then, one can make efficient forecasts of a single variable benefiting from the information of a large number of variables (see Luciani (2017) (Chapter 17)).

- the mixed frequency data models in state space form where the low-frequency variable is represented as a high-frequency variable with missing data. The missing data is then estimated using for instance a Kalman filter and the low-frequency variable can be forecasted.

- temporal disaggregation techniques combined with a single or multi-factor composite indicator such as in Frale et al. (2008) and Frale et al. (2011).

The fact that the dataset considered often include mixed frequency data has emphasized, especially in the last years, the growing importance of various types of mixed frequency modelling associated almost to all main statistical models. Finally, it is not uncommon that some models, especially dynamic factor and principal components models, are combined with other statistical and econometric techniques, such as temporal disaggregation, VAR models, unobserved component models, etc.

15.5 Review of the literature

When forecasting key macroeconomic variables, it is important to consider a reasonable number of regressors to avoid any estimation problem. As seen in the previous section, there are two main approaches to get this small number of regressors: (i) using a small set of carefully chosen predictors, or (ii) considering a large set of predictors whose information is then summarized in a small set of synthetic variables. The selection of variables in the first approach relies either on experts' experience and intuition, or on quantitative approaches such as the LARS algorithm introduced by Efron et al. (2004) and used in Chapter 18. The second approach relies on variable reduction methods such as the partial least squares and Bayesian shrinkage regression presented in Chapter 5 and used in Chapters 5 and 16.

Regarding the models used, the first approach is to consider the usual linear models (VAR, VECM, etc.). For instance, Pirschel and Wolters (2014) use Bayesian vector autoregression models to shrink large datasets of dependent variables, and Bayesian moving averaging techniques to aggregate the forecasts of a large number of small VAR models. The gain of these methods over an AR model is modest and none could predict the great recession in the German GDP growth. The main alternatives to these classical models are the dynamic factor models and the mixed-frequency data models. The dynamic factor models were first proposed by Geweke (1977) and gained popularity with Sargent and Sims (1977) and Stock and Watson (1989). They consider that few latent factors drive a large number of the economic variables. The information of this large number of variables is thus summerized in the latent factors estimated through state-space models. We refer to Stock and Watson (2010) for a review of dynamic factor models. The mixed data frequency models include the mixed data sampling models (MIDAS), the bridge models and state space models. The mixed data sampling models were first proposed by Ghysels et al. (2004).

First focusing on financial applications, this method has since been used to forecast macroeconomic variable, see Clements and Galvao (2008) and Clements and Galvao (2009) among others. The bridge equation, introduced in Baffigi et al. (2004), relate low-frequency variables to high-frequency time aggregated variables. Specific high frequency time series models provide forecasts of the high frequency variables which are aggregated and inserted in the bridge equations to obtain the forecast of the low frequency variable. Finally, the state-space representation of a model, where the low-frequency variable is represented as a high-frequency

variable with missing data, enables to estimate the missing data and to forecast the low-frequency variable. In particular, Mariano and Murasawa (2010) introduce the MF-VAR to now/forecast GDP growth, in small scale applications. Giannone et al. (2008) and Banbura and Rünstler (2011) applied the methodology to large scale models. We refer to Foroni and Marcellino (2013) for a recent review of the mixed-frequency data models. An interesting and quite innovative applied paper comparing the number of mixed-frequency models is presented by Aprigliano et al. (2017). In this paper the authors try to compile a daily updated indicator of the euro area GDP quarterly growth by exploiting a number of daily variables using several mixed-frequency modelling techniques.

15.6 Some regularly available coincident and leading indicators

Several coincident or leading indicators are regularly produced and disseminated by public and private institutions. Beside cyclical and turning points indicators, Eurostat disseminates the following indicators in its monthly publication, Eurostatistics:[1]

- €-coin: a real-time monthly estimate of the euro area GDP growth, produced by Banca d'Italia. It focuses on the medium to long-run component of GDP and relies on principal components for variable reduction and on a low-pass filtering. See Altissimo et al. (2007) for technical details.[2]

- Euroframe: a leading monthly estimate of the euro area GDP growth, produced by a network of ten research institutes (CPB, DIW Berlin, ESRI, ETLA, IFW, NIESR, OFCE, Prometeia, WIFO and CASE). It aims to anticipate the euro area GDP growth 2 quarters ahead of official statistics.[3]

In addition to the dissemination of existing indicators, Eurostat experiments with its own GDP coincident indicator compilation whose development is presented in Charpin (2009), Charpin (2011), Charpin (2017).

A comparison of the above-mentioned indicators is presented in Mazzi and Ruggeri (Chapter 21).

Among the other indicators regularly published, we note:

- the Now-Casting Index: a monthly nowcasting indicator of euro area GDP growth, among other countries and series. It is produced by Now-Casting Economics Ltd, founded by Lucrezia Reichlin. This indicator is based on dynamic factor models and variable reduction techniques

- the Chicago Fed National Activity Index: a monthly coincident index of the economic activity for the US, based on Stock and Watson (1999) and produced by the Federal Reserve Bank of Chicago.

At the institutional level for the euro area we would like to mention the leading GDP indicators produced by the Statistical offices of France and Italy, and by the ECB. Among the non-euro area countries, without going into many details especially for the US where the number of available indicators is quite high, we would like to mention the Indian composite indicator (see Rajeswari (2010)) as well as the efforts in compiling monthly GDP estimates made by several countries such as Peru (see Garcia (2010)) and Chile (see Pedersen (2013)).

15.7 Guide to the part

This part of the handbook proposes chapters focusing on the main approaches presented above in the forecasting of macroeconomic variables. Chapter 16 compares a dynamic factor model in the case of small and

---

[1]See http://ec.europa.eu/eurostat/web/euro-indicators/publications/official-publications
[2]See http://eurocoin.cepr.org/
[3]See http://www.euroframe.org/Indicator.html

large data sets and with the addition of different variable reduction methods. Chapter 17 reviews the literature on large scale dynamic factor models in forecasting. Chapter 18 presents a coincident estimate of the quarterly GDP growth rate based on the LARS variable selection method and a bridge error correction model.

The aim of Chapter 16, Camba-Mendez et al. (2017), is to assess and compare the forecasting ability of the automatic leading indicator (ALI) introduced in Camba-Mendez et al. (2001), when it is enhanced with different variable reduction methods and the case of either small or large datasets. This indicator forecasts the growth rate of the industrial production. It is based on a small scale dynamic factor model whose estimation is composed of two steps. First, the factors are extracted. Second, the factors are included in the model which is estimated and used for forecasting. The variable reduction methods considered are the partial least squares (PLS), extracting either 1 or 3 factors, and the Bayesian shrinkage regression with N/2 and 2N as shrinkage parameters, where N is the number of predictors in the dataset. These reduction methods are introduced in Kapetanios et al. (2017) (Chapter 6). With a small dataset of predictors and across different regions (euro are, France, Germany and Italy), the addition of reduction methods does not outperform the simple ALI with no reduction method. With a large dataset, the results are more mitigated. Overall, the reduction methods provided better forecasts in the short-term (one step-ahead).

Chapter 17, Luciani (2017), reviews the literature on large dimensional dynamic factor models with a focus on the nowcasting of macroeconomic variables. In its methodological review, it underlines the challenges implied by non-synchronous releases of data and mixed frequencies. It also analyses the performance of two indicators: the Eurocoin indicator and the Chicago Fed National Activity Index, both presented above.

Chapter 18, Charpin (2017) presents the development of a coincident quarterly growth rate of GDP, using monthly data. The dataset considered is composed of survey data (soft data), macroeconomic variables (hard data) and financial data. They are all considered at different lags and the survey data is also considered both in levels and in variations. From this large dataset, between 12 and 15 variables are selected using the least angle regression (LARS) algorithm. Then, few synthetic orthogonal variables are obtained by principal components analysis. The resulting factors enter in a bridge error correction model of the GDP growth rate and the estimation is obtain. The accuracy of the estimations are at least as good as those from more sophisticated models such as the dynamic factor models presented in Chapter 17.

15.8 Conclusions

This chapter has introduced the main quantitative approaches based on composite indicators used in forecasting key macroeconomic variables in the short-term. As seen in Section 15.6, the large variety of methods implied by this still ongoing research have each their pros and cons. The most sophisticated model do not always outperform the more simple ones as observed in the following chapters. The latest developments attempt to combine these various methods to mitigate the cons and provide more accurate forecasts.

Bibliography

Altissimo F., Cristadoro R., Forni M., Lippi M. and Veronese G. (2007). New Eurocoin: Tracking economic growth in real time, The Review of Economics and Statistics, No. 92, 1024-1034

Aprigliano V., Foroni C., Marcellino M., Mazzi GL. and Venditti F. (2017), A daily indicator of economic growth for the euro area, International Journal of Computational Economics and Econometrics, 7, (1), 43-63.

Baffigi A., Golinelli R. and Parigi G. (2004). Bridge models to forecast the Euro area GDP, International Journal of Forecasting, 20(3), 447-460.

Banbura M. and Rünstler G. (2011). A Look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP, International Journal of Forecasting, 27(2), 333-346.

Camba-Mendez G., Kapetanios G, Papailias F. and Weale M. (2017). Chapter 16: An Automatic Leading Indicator with variable reduction using small and large datasets: forecasting the industrial production growth for euro area economies, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Camba-Mendez G., Kapetanios G, Smith R.J. and Weale M. (2001). An automatic leading indicator of economic activity: forecasting GDP growth for European countries, Econometrics Journal, 4, S56-S90

Charpin F. (2009). Estimation précoce de la croissance: de la régression LARS au modèle à facteurs. Revue de l'OFCE no. 108:31-48.

Charpin F. (2011). Réévaluation des modèles d'estimation précoce de la croissance. Revue de l'OFCE no. 118:129-142.

Charpin F. and Mazzi G.L. (2017). Chapter 13: An empirical investigation of combining forecasting techniques, in: Handbook on Rapid Estimates, ed. by D. Ladiray and G.L. Mazzi.

Charpin F. (2017). Chapter 18: Regression, principal components and small scale factors based models, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Clements M.P. and Galvao A.B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting US output growth, Journal of Business and Economic Statistics, 26, 546-554.

Clements M.P. and Galvao A.B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models, Journal of Applied Econometrics, 24(7), 1057-1217.

Efron B., Hastie T., Johnstone T. and Tibshirani R. (2004). Least Angle Regression, Anals of statistics, 32 (2), 407-499

Foroni C. and Marcellino M. (2013). A survey of econometric methods for mixed-frequency data, Norges Bank Research, Working Paper 2013/06.

Frale C., Marcellino M., Mazzi G.L. and Proietti T. (2008) A monthly indicator of the Euro Area GDP, CEPR Discussion Paper DP7007

Frale C., Marcellino M., Mazzi GL. and Proietti T. (2010), Survey Data as Coincident or Leading Indicators, Journal of Forecasting, 29, 109-131.

Frale C., Marcellino M., Mazzi G.L. and Proietti T. (2011), EUROMIND: a monthly indicator of the euro area economic conditions, Journal of the Royal Statistical Society, series A, 174, part 2, 439-470.

Garcia-Zanabria J.A. (2010), Monthly estimate of GDP in Peru, Presentation at the 6th collquium on modern tools for business cycle analysis: the lessons from the global economic crisis, Eurostat, September 26-29 2010, Luxembourg

Geweke J. (1977). The Dynamic Factor Analysis of Economic Time Series, in Latent Variables in Socio-Economic Models, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Giannone D., Reichlin L. and Small D. (2008). Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases, Journal of Monetary Economics, 55, 665-676.

Ghysels E., Santa-Clara P. and Valkanov R. (2004). The MIDAS touch: MIxed DAta Sampling regression models, mimeo, Chapel Hill, N.C.

Kapetanios G., Marcellino M. and Papailias F.(2017). Chapter 6: Variable Reduction and Variable Selection Methods using Small, Medium and Large Datasets: A Forecast Comparison for the PEEIs, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Luciani M. (2017). Chapter 17: Large Dimensional Dynamic Factor models in real-time, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mariano R. and Murasawa Y. (2010). A coincident index, common factors, and monthly real GDP, Oxford Bulletin of Economics and Statistics, 72(1), 27-46.

Pedersen M. (2013). Extracting GDP signals from the monthly indicator of economic activity: Evidence from Chilean real-time data. OECD Journal: Journal of Business Cycle Measurement and Analysis, OECD Publishing, Centre for International Research on Economic Tendency Surveys, vol. 2013(1), pages 1-16.

Pirschel I. and Wolters M.H. (2014). Forecasting German Key Macroeconomic Variables Using Large Dataset Methods. Manuscript, University of Kiel.

Rajeswari T. (2010). Construction of a Composite Leading Indicator for India. Third International Seminar on Early Warning and Business Cycle Indicators, Nov. 2010.

Sargent T.J. and Sims C.A. (1977). Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory, in New Methods in Business Cycle Research, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.

Stock J.H. and Watson M.W. (1989). New Indexes of Coincident and Leading Economic Indicators, NBER Macroeconomics Annual 1989, 351-393.

Stock J.H. and Watson M.W. (1999). Forecasting inflation, Journal of Monetary Economics, 44 (2), 293-335

Stock J.H. and Watson M.W. (2010). Dynamic Factor Models, in Oxford Handbook of Economic Forecasting, Michael P. Clements and David F. Hendry (eds), Oxford University Press.

16 Leading Indicator, Variable Reduction and Selection Methods

Authors

Gonzalo Camba-Mendez
ECB
gonzalo.camba-mendez@ecb.europa.eu

Martin R. Weale
King's College, University of London
martin.weale@kcl.ac.uk

George Kapetanios
King's College, University of London
george.kapetanios@kcl.ac.uk

Fotis Papailias
King's College, University of London and quantf research
f.papailias@kcl.ac.uk

Table of Contents

16.1 Introduction

The issue of forecasting key macroeconomic variables has been constantly on the debate over the past years. In general, methodologies can be divided in two broad categories characterised by the size of the set of explanatory variables taken into consideration. Stock and Watson (1991), Camba-Mendez et al. (2001), Aruoba et al. (2009), Aruoba and Diebold (2010) and Camacho and Perez-Quiros (2010) are some examples where a small number of wisely selected predictors is used. Then, under the assumption of non cross-correlated errors the factor models are estimated by maximum likelihood using the Kalman filter.

On the other hand, the seminal work by Stock and Watson (2002a) suggests forecasting using a large set of predictors where information is summarised using principal components estimation. The forecasting performance of other variable reduction methods have been recently studied by Del Mol et al. (2006), Kapetanios et al. (2014a) among others and include the Bayesian shrinkage regression, partial least squares and others.

A difficult issue in both the above categories is the choice of the set of predictors and the actual variables. Boivin and Ng (2006) suggest that cross-correlation of regressors in large datasets might result in inaccurate forecasts and hence a smaller set is more likely to provide a smaller average forecast error. A key to this problem is the use of variable selection methods. Kapetanios (2007) and Kapetanios et al. (2014a) use non-standard optimisation of information criteria in order to identify the appropriate instruments and they forecast EU macroeconomic variables with encouraging results. A recent study by Alvarez et al. (2012) also addresses this issue and compares small and large scale dynamic factor models in the Stock and Watson dataset for the US. However, their approach is limited in that particular class of models.

The purpose of this chapter is to evaluate and compare the forecasting ability of an Automatic Leading Indicator and standard variable reduction methods using small and large datasets. We predict the growth rate of the industrial production of the Euro Area (16), France, Germany and Italy. The variable reduction methodologies include: the Bayesian shrinkage regression and the partial least squares. The Automatic Leading Indicator (ALI) model is constructed in the spirit of Camba-Mendez et al. (2001) which belongs to the category of small scale dynamic factor models; see Chapter 13 for large-scale models. Composite Indicators (as defined in Chapter 2) can be constructed using the hereby reviewed methodologies.

The rest of the chapter is organised as follows: Section 16.2 briefly describes the methodology, Section 16.3 is concerned with the forecasting algorithm, Section 16.4 discusses the results and Section 16.5 summarises the conclusions.

16.2 Methodology

Similar to Chapter 5, we consider the following regression model,

$$y_t = \alpha + \beta^{0'} x_t^0 + \epsilon_t, \quad t = 1, \ldots, T, \tag{16.1}$$

where x_t^0 is a k-dimensional vector of stationary predetermined variables. The superscript ^{0} denotes the true regression model. Let the set of all available variables at time t be represented by the N-dimensional vector $x_t = (x_{1,t}, \ldots x_{N,t})'$, where it is currently assumed that the set of variables in x_t^0 is also contained in x_t. The aim of the analysis is to determine x_t^0.

The automatic leading indicator (ALI) model is introduced by Camba-Mendez et al. (2001) and is a small scale dynamic factor model which includes a two-step procedure. In the first step the factors are extracted and in the second stage a VAR model is estimated and used for forecasting. Consider the following model for the N vector of exogenous variables X,

$$X_t = Bs_t + u_t,$$
$$C\left(L\right)s_t = \eta_t, \tag{16.2}$$

where B is an (n,k) matrix of unknown parameters, s_t is a k vector of factors that follows a stationary AR(p) process with disturbances η and u_t is an N vector of disturbances. The estimation of the unknown parameters in (16.2) and the extraction of factors may be combined in the following step-wise fashion. Given knowledge of B, $C(L)$ and the variance matrices of u_t and η_t, (16.2) can be written in state space form with the Kalman filter used to extract s_t from observation on x_t. Secondly, given the factors s_t, the parameter matrices B, $C(L)$ and the variance matrices of u_t and η_t may be estimated by quasi-maximum likelihood. This step-wise procedure may be iterated until convergence; see Harvey (1993) for further details.

The factors s_t obtained above are then incorporated into a VAR model to forecast y_t as follows,

$$A_y(L)y_t = As(L)s_t + \epsilon_t, \tag{16.3}$$

where ϵ_t is a zero-mean conditionally homoskedastic and serially uncorrelated error process with positive definite variance matrix uncorrelated with the error processes u_t and η_t. Then given the estimated factors and parameters the model in (16.3) is estimated via OLS (or maximum likelihood) and the parameter estimates are used in the forecasting exercise as described in the next section. The lags in the AR and VAR model and the choice of factors are determined by a series of tests that include serial correlation and Granger causality; see Camba-Mendez et al. (2001) for more details. It is important to notice here that the ALI is always used with the small set of predictors (see Table 16.A.9).

See Chapter 5 for a detailed explanation of Partial Least Squares (PLS) and Bayesian Shrinkage Regression (BR).

16.3 Forecasting and Data Description

We perform a forecasting exercise using the projection method as described in Stock and Watson (2002a). This method, also known as direct approach, is more robust in the presence of possible model mis-specification. The forecasts are given by,

$$\widehat{y}_{t+h/T}^f = \widehat{\beta}^{h\prime} z_{t/T}, \tag{16.1}$$

where $\widehat{\beta}^h$ is obtained by regressing y_t on z_{t-h} and h denotes the forecast horizon. z_t is a k-dimensional vector of variables and can be equal to x_t or to k-factors series depending on the choice of the estimation method. In the case of the ALI z_t might include the autoregressive components of y_t as well.

At first, we set the $\max steps$ ahead, h. Then, we specify the evaluation period, $Eval$, and we omit h observations completely out of the sample. This allows us to end up with a number of $Eval$ forecasts for any given step h. A summary of the pseudo out-of-sample forecasting algorithm follows.

1. Use an initial sample of T_1 observations ($T_1 = T - Eval - h$),

2. With any method described in this section obtain x'_t, $t = 1, 2, ..., T_1$,

3. For $j = 1, 2, ..., h$ steps regress y_t on z'_{t-h} and obtain $\widehat{\beta}^h = \left(\widehat{\beta}^1, ..., \widehat{\beta}^h\right)'$,

4. Calculate the forecasts of $\widehat{y}_{t+h}^f$ using z'_t and $\widehat{\beta}^h$, hence $\widehat{y}^f = \left(\widehat{y}_1^f, ..., \widehat{y}_h^f\right)'$,

5. We repeat the whole procedure increasing the initial sample T_1 to $T_l = T_{l-1} + 1$ until $T_l = T - h$.

At the end of this process we have gathered a number of $Eval$ forecast values for any step h. We consider two looping procedures: (i) a recursive looping where the initial sample T_1 is augmented by one observation at each time and (ii) a rolling approach where a fixed sample of T_1 length moves across time.

The forecast error is then calculated as,

$$\widehat{e}^f_{t+h} = y_{t+h} - \widehat{y}^f_{t+h}, \tag{16.2}$$

and the statistics of interest can be computed. We are particularly interested in the Root Mean Squared Forecast Error[1] defined as,

$$RMSFE_h = \sqrt{\frac{1}{Eval} \sum_{j=1}^{Eval} \left(e^f_{t+h,j}\right)^2}. \tag{16.3}$$

The small dataset of predictors is described in Table 16.A.9 and consists of various government bond spreads, the real effective exchange rate, house lending rates, stock market return, money supply and the survey-based economic sentiment indicator[2] for each economy. The dependent variable subject to forecasting is the growth rate of the industrial production of the Euro Area (16), France, Germany and Italy. The data was collected using Macrobond Financial and the dates span from Jan. 1996 to Jan. 2009.

The large dataset of predictors consists of 195 monthly variables (source: European Commission, Eurostat, PEEIs, the Eurostat labels can be found in the Table 16.B.1) spanning from Jan. 1996 to Jan. 2009. The dataset is the same used in Foroni and Marcellino (2011) and it contains a large universe of variables that are potentially useful instruments in forecasting key macroeconomic variables in the Euro Area. Furthermore, in the spirit of Stock and Watson (2002a) we have transformed the series for stationarity using first differences or $\log$ differences appropriately (although notice in Table 16.C.1 some of the variables remained unchanged). Hence, the resulting data used in the forecast exercise contains growth rates from Feb. 1996 to Jan. 2009 (inclusive).

It is important to notice here that both small and large datasets of predictors are correctly date-aligned and the forecasting exercise uses the same time period in all experiments. The cross-validation (forecasting evaluation) period is set to 84 months, starts in Feb. 2002 and ends in Jan. 2009. The forecast horizon is set to $h = 12$ months.

We normalise the regressors to zero mean and unit variance series in all cases.

---

[1]Diebold-Mariano Statistics are available on request from the authors.
[2]Source: European Commission.

16.4 Discussion of Results

16.4.1 Small Dataset of Predictors

Tables 16.A.1-16.A.4 present the RMSFE of the benchmark ALI model that uses the small set of predictors and the relative RMSFE of PLS and BR models using the same dataset.

In Table 16.A.1 we present the forecasting exercise for the growth rate of the industrial production for the Euro Area economy. We see that the average RMSFE of the ALI model across all forecast horizons is equal to 0.009. In the first step ahead none of the other variable reduction methods performs better compared to the benchmark. PLS with three factors is slightly worse with a relative RMSFE equal to 1.02 and the Bayesian regression with shrinkage parameter $v = 0.5N$ presents the highest forecast error with relative RMSFE equal to 1.073. The same can be said for all subsequent steps up to $h = 7$ and for $h = \{9, 10\}$. In $h = 8$ we have the PLS(3) and the BR(2N) being relatively better than the ALI with relative RMSFE equal to 0.984 and 0.987 respectively. BR(2N) provides better forecasts compared to the benchmark model in steps $h = \{11, 12\}$ as well with relative RMSFE equal to 0.992 and 0.989 respectively. PLS(1) is slightly better than the ALI in step $h = 11$ and BR(0.5N) in step $h = 12$ with relative RMSFE of 0.997 and 0.998 respectively.

The forecasting results for the growth rate of the industrial production for France can be found in Table 16.A.2. As in the EA case, the ALI should be preferred compared to the other variable reduction models. In steps $h = \{1...4, 9, 10\}$ ALI is always better as the relative RMSFE of all other methods is constantly above unity. For 5 to 7 steps ahead PLS(1) forecasts are slightly better compared to the benchmark's with a relative RMSFE of 0.999 and in step 8 BR(2N) presents a relative RMSFE equal to 0.996. In steps 11 and 12 both BR models are better.

Table 16.A.3 describes the forecasting results for Germany's industrial production growth rate. In steps $h = \{1, 3\}$ PLS(3) performs better with a RMSFE of 0.955 and 0.985 relative to the ALI benchmark. In steps $h = \{2, 4...8\}$ ALI outperforms all other models with an actual RMSFE of 0.012 across all steps. However in steps $h = \{9, 11, 12\}$ the other methods provide more accurate forecasts. In step $h = 12$ the BR(0.5N) has a relative RMSFE of 0.975 and the BR(2N) a relative RMSFE of 0.97.

The last case is that of Italy in Table 16.A.4. This is another example (as in the case of EA) where ALI is shown to outperform all other methods. In steps $h = \{3, 8, 11\}$ PLS(1), BR(0.5N) and BR(2N) perform slightly better where in the best case the relative RMSFE is 0.994.

16.4.2 Large Dataset of Predictors

We continue our study using the variable selection and variable reduction methods as described in Section 16.2 of the chapter with a large dataset of predictors. The benchmark model is still the ALI when using the small dataset.

For the case of the EA we see in Table 16.A.5 that in the first step PLS(1), PLS(3) and BR(2N) provide better forecasts compared to the ALI. In particular, the actual RMSFE of the ALI is 0.008 and the relative RMSFE of the above models is 0.964, 0.914 and 0.952 respectively. For 2, 4 and 8 steps ahead the ALI outperforms all other competing methods with an average RMSFE of 0.008. PLS(1) and PLS(3) seem to be the most robust alternatives to the benchmark as they are characterised with relative RMSFE smaller than unity for 1, 3, and 5 to 12 steps ahead. ALI loses its forecasting power as we move into future steps ahead. Especially, for $h = 12$ ALI has an actual RMSFE of 0.01 and PLS(1) has a relative RMSFE of 0.882.

The ALI performs better compared to the other methods for 2 to 4 and 6 to 8 steps ahead when we forecast the industrial production growth for France. In Table 16.A.6 we see that almost all methods outperform the ALI in step 1. The better forecasts are provided by SA_{BIC} with a relative RMSFE equal to 0.901. PLS(1),

PLS(3) and BR(2N) provide forecasts with smaller error for 1, 5 and 9 to 12 steps ahead. In this case, the ALI performs relatively better compared to the case of the EA as we move to future forecasts. For 12 steps ahead we see that the benchmark's actual RMSFE is equal to 0.015 and its best competitors are the BR(2N) and BR(0.5N) with 0.963 and 0.981 relative RSMFE.

Table 16.A.7 presents the forecasting exercise for the growth rate of the industrial production for Germany. The qualitative conclusion here is similar to that of France. As before, the ALI provides better forecasts for 2 and 4 to 11 steps ahead (on average). The forecast with the smallest error is given by PLS(3) with a relative RMSFE equal to 0.946. A similar result is given in step 3, where almost all methods are better than the benchmark. Now, BR(2N) has the smallest forecast errors with relative RMSFE equal to 0.892 and 0.906 respectively. For 4 to 10 steps ahead ALI seems to provide good forecasts on average apart from step 5 where PLS(1) has a relative RMSFE of 0.98. For the last step ahead, $h = 12$, ALI's RSMFE is equal to 0.014 and BR(2N) has the smallest relative RMSFE of 0.961.

Moving to last set of empirical results for the case of the industrial production growth for Italy we see again that the ALI performs better in 2 to 9 steps ahead. Table 16.A.8 shows that in the first step PLS(1), PLS(3) and BR(2N) return forecasts with smaller error on average. ALI has an actual RMSFE of 0.014 and PLS(3) has a relative RMSFE of 0.939. As h increases the ALI returns forecasts with larger errors on average but a large differences can be seen for $h = 11$ and $h = 12$ steps ahead. PLS(1) and PLS(3) have a relative RMSFE 0.97 and 0.917 respectively.

16.5 Concluding Remarks

In this chapter we approach the issue of forecasting macroeconomic variables using indicators from small and large datasets. The methods we employ include an Automatic Leading Indicator model of Camba-Mendez et al. (2001), which belongs to the class of the small scale dynamic factor models, and standard variable reduction methods.

Our overall work indicates that among the variable reduction methods using a small set of predictors, the ALI is more likely to provide better forecasts on average. However, the ALI performs well only in the "medium-term" forecasts (2-10 steps ahead) using a large dataset of predictors. The variable reduction methodologies, and especially the PLS model with 1 factor, are more likely to result in forecasts with smaller errors in the first and 8 to 12 steps ahead.

To conclude with, our empirical approach suggests that the models described here should, at least, be considered by researchers interested in model averaging methodologies for forecasting key European macroeconomic variables. Furthermore, we show that the use of out-of-sample forecasting using variable selection and variable reduction methodologies might be of importance to the applied research who focuses on the construction of (cyclical) composite indicators.

Annex

16.A Tables

Table 16.A.1: Forecasting the industrial production growth rate of the EA using a small set of predictors

					EA Industrial Production Growth Rate							
					Forecast Horizon							
h	1	2	3	4	5	6	7	8	9	10	11	12
					RMSFE							
ALI	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.009	0.010
					Relative (to the ALI) RMSFE							
PLS(1)	1.024	1.035	1.013	1.048	1.026	1.014	1.020	1.010	1.029	1.019	0.997	1.000
PLS(3)	1.020	1.051	1.036	1.091	1.022	1.038	1.032	0.984	1.050	1.032	1.012	1.048
BR(0.5N)	1.073	1.068	1.018	1.088	1.043	1.037	1.031	1.025	1.046	1.045	1.005	0.998
BR(2N)	1.048	1.051	1.006	1.059	1.017	1.017	1.004	0.987	1.024	1.014	0.992	0.989
					Cross-validation: 84 periods							

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.2: Forecasting the industrial production growth rate of France using a small set of predictors

					France Industrial Production Growth Rate							
					Forecast Horizon							
h	1	2	3	4	5	6	7	8	9	10	11	12
					RMSFE							
ALI	0.013	0.013	0.013	0.013	0.014	0.014	0.014	0.014	0.014	0.014	0.015	0.015
					Relative (to the ALI) RMSFE							
PLS(1)	1.008	1.019	1.018	1.003	0.999	0.999	0.999	1.008	1.007	1.018	1.001	1.001
PLS(3)	1.009	1.020	1.025	1.009	1.020	1.002	1.024	1.005	1.006	1.037	1.004	0.991
BR(0.5N)	1.064	1.022	1.020	1.049	1.032	1.011	1.022	1.007	1.045	1.006	0.993	0.985
BR(2N)	1.045	1.015	1.013	1.031	1.015	1.002	1.010	0.996	1.021	1.000	0.991	0.981
					Cross-validation: 84 periods							

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.3: Forecasting the industrial production growth rate of Germany using a small set of predictors

Germany Industrial Production Growth Rate											
Forecast Horizon											
h	1	2	3	4	5	6	7	8	9	10	11	12
RMSFE											
ALI	0.013	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.013	0.014
Relative (to the ALI) RMSFE											
PLS(1)	1.010	1.023	0.990	1.025	1.008	1.007	1.007	1.001	0.995	1.059	0.991	0.991
PLS(3)	0.955	1.024	0.985	1.036	1.056	1.070	1.041	1.012	1.018	1.070	1.033	0.982
BR(0.5N)	1.067	1.029	1.032	1.025	1.053	1.104	1.048	1.033	1.054	1.051	1.016	0.975
BR(2N)	1.034	1.015	1.006	1.017	1.031	1.052	1.024	1.012	1.026	1.020	0.996	0.970
Cross-validation: 84 periods											

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.4: Forecasting the industrial production growth rate of Italy using a small set of predictors

Italy Industrial Production Growth Rate											
Forecast Horizon											
h	1	2	3	4	5	6	7	8	9	10	11	12
RMSFE											
ALI	0.014	0.014	0.014	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.016	0.016
Relative (to the ALI) RMSFE											
PLS(1)	1.006	1.000	0.999	1.002	1.003	1.003	1.003	1.001	1.003	1.006	0.997	1.003
PLS(3)	1.055	1.018	1.001	1.013	1.017	1.009	1.029	1.000	1.005	1.019	1.013	1.021
BR(0.5N)	1.039	1.014	0.994	1.013	1.015	1.003	1.014	0.994	1.000	1.015	1.013	1.004
BR(2N)	1.025	1.009	0.992	1.008	1.012	1.002	1.011	0.997	1.001	1.011	1.008	1.006
Cross-validation: 84 periods											

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.5: Forecasting the industrial production growth rate of the EA using a large set of predictors

EA Industrial Production Growth Rate											
Forecast Horizon											
h	1	2	3	4	5	6	7	8	9	10	11	12
RMSFE											
ALI	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.009	0.010
Relative (to the ALI) RMSFE											
$PLS(1)$	0.964	1.007	0.936	1.024	0.980	0.994	0.979	1.000	0.983	0.992	0.911	0.906
$PLS(3)$	0.914	1.007	0.941	1.034	0.986	1.003	0.981	0.995	0.983	0.987	0.909	0.882
$BR(0.5N)$	1.034	1.163	1.124	1.285	1.154	1.290	1.162	1.195	1.184	1.192	1.089	0.967
$BR(2N)$	0.952	1.052	0.995	1.119	1.030	1.085	1.029	1.067	1.036	1.032	0.937	0.879
Cross-validation: 84 periods											

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.6: Forecasting the industrial production growth rate of France using a large set of predictors

France Industrial Production Growth Rate											
Forecast Horizon											
h	1	2	3	4	5	6	7	8	9	10	11	12
RMSFE											
ALI	0.013	0.013	0.013	0.013	0.014	0.014	0.014	0.014	0.014	0.014	0.015	0.015
Relative (to the ALI) RMSFE											
PLS(1)	0.979	1.014	0.999	0.999	0.954	1.029	1.008	1.005	0.993	0.987	0.977	0.998
PLS(3)	0.942	1.019	1.002	1.005	0.960	1.038	1.016	1.006	0.985	0.973	0.964	0.990
BR(0.5N)	0.995	1.014	1.069	1.169	1.039	1.128	1.120	1.087	1.068	1.046	1.052	0.981
BR(2N)	0.969	0.989	1.016	1.059	0.972	1.043	1.050	1.035	1.003	0.985	0.990	0.963
Cross-validation: 84 periods											

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.7: Forecasting the industrial production growth rate of Germany using a large set of predictors

h	1	2	3	4	5	6	7	8	9	10	11	12
						Germany Industrial Production Growth Rate						

Germany Industrial Production Growth Rate
Forecast Horizon

h	1	2	3	4	5	6	7	8	9	10	11	12
						RMSFE						
ALI	0.013	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.013	0.014
						Relative (to the ALI) RMSFE						
PLS(1)	0.979	1.004	0.916	0.994	1.005	0.984	1.018	1.014	1.005	0.997	0.998	1.008
PLS(3)	0.946	1.007	0.926	1.001	1.005	0.993	1.026	1.020	1.018	0.994	1.015	0.990
BR(0.5N)	1.065	1.136	0.961	1.134	1.039	1.162	1.096	1.077	1.147	1.106	1.061	0.993
BR(2N)	0.983	1.042	0.906	1.032	1.004	1.032	1.031	1.018	1.041	1.022	0.978	0.961

Cross-validation: 84 periods

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.8: Forecasting the industrial production growth rate of Italy using a large set of predictors

Italy Industrial Production Growth Rate
Forecast Horizon

h	1	2	3	4	5	6	7	8	9	10	11	12
						RMSFE						
ALI	0.014	0.014	0.014	0.014	0.015	0.015	0.015	0.015	0.015	0.015	0.016	0.016
						Relative (to the ALI) RMSFE						
PLS(1)	0.961	1.014	0.997	1.016	0.989	1.003	0.997	0.986	1.024	0.984	0.970	0.954
PLS(3)	0.939	1.022	1.005	1.029	0.996	1.008	1.005	0.999	1.026	0.981	0.984	0.917
BR(0.5N)	1.001	1.082	1.129	1.137	1.049	1.050	1.028	1.084	1.070	0.994	0.993	0.954
BR(2N)	0.968	1.033	1.050	1.065	1.010	1.010	1.000	1.024	1.033	0.971	0.968	0.944

Cross-validation: 84 periods

Note: **ALI** denotes the Automatic Leading Indicator model with 1 factor, **PLS(1)** denotes the Partial Least Squares with 1 factor, **PLS(3)** denotes the Partial Least Squares with 3 factors, **BR(0.5N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 0.5N$, **BR(2N))** denotes the Bayesian Shrinkage regression with shrinkage parameter $v = 2N$.

Table 16.A.9: Small dataset of predictors

Small Dataset of Predictors		
Variable	Description	Transformation
Rates		
1 10yr Government Benchmark Bond Yield	Government Benchmarks	Level
2 10yr Govt Bond/3m US Govt Bond Spread	Government Benchmarks	Level
3 10yr Govt Bond/10yr US Govt Bond Spread	Government Benchmarks	Level
4 2yr Govt Bond/2yr US Govt Bond Spread	Government Benchmarks	Level
5 Corporate Bonds Yield	Corporate Benchmarks	Level
6 Real Effective Exchange Rate	FX Indices, BIS	Level
7 House Lending	Lending for House Purchase	Level
8 Local Equity Index Return	Equity Indices	Growth
9 Local Volatility Index Return	Volatility Indices	Growth
10 Money Supply: M3	Monetary Aggregates	Growth
11 Economic Sentiment Indicator	Economic Sentiment Surveys	Level
Dependent Variable		
1 Industrial Production (SA) Growth	Industrial Production Index	Growth

16.B Large dataset of predictors: Labels

Labels and Groups of Variables used

#	Label	#	Label
1	CP-HI00XEFU	101	1000-PERS-LM-UN-T-TOT
2	CP-HI00XEF	102	IS-IP
3	CP-HI00XES	103	IS-IP-F-CC1
4	CP-HI00XE	104	IS-IP-F-CC2
5	CP-HI00XTB	105	IS-IP-F
6	CP-HI00	106	B-D-IS-IP
7	CP-HI01	107	B-C-IS-IP
8	CP-HI02	108	B-IS-IP
9	CP-HI03	109	C-IS-IP
10	CP-HI04	110	C-ORD-IS-IP
11	CP-HI05	111	D-IS-IP
12	CP-HI06	112	MIG-CAG-IS-IP
13	CP-HI07	113	MIG-COG-IS-IP
14	CP-HI08	114	MIG-DCOG-IS-IP
15	CP-HI09	115	MIG-ING-IS-IP
16	CP-HI10	116	MIG-NDCOG-IS-IP
17	CP-HI11	117	IS-EPI-F
18	CP-HI12	118	B-D-IS-EPI
19	CP-HIE	119	B-E36-IS-EPI
20	CP-HIF	120	B-C-IS-EPI
21	B-D-IS-PPI	121	B-IS-EPI
22	B-E36-IS-PPI	122	B-TO-E36-IS-EPI
23	B-C-D-IS-PPI	123	C-IS-EPI
24	B-IS-PPI	124	D35-E36-IS-EPI
25	B-TO-E36-IS-PPI	125	D-IS-EPI
26	C-IS-PPI	126	E36-IS-EPI
27	C-ORD-IS-PPI	127	E-IS-EPI
28	D-IS-PPI	128	MIG-CAG-IS-EPI
29	E36-IS-PPI	129	MIG-COG-IS-EPI
30	MIG-CAG-IS-PPI	130	MIG-DCOG-IS-EPI
31	MIG-COG-IS-PPI	131	MIG-ING-IS-EPI
32	MIG-DCOG-IS-PPI	132	MIG-NDCOG-IS-EPI
33	MIG-ING-IS-PPI	133	MIG-NRG-IS-EPI
34	MIG-NDCOG-IS-PPI	134	G45-IS-EPI
35	MIG-NRG-IS-PPI	135	B-C-IS-ITD
36	D35-E36-IS-IMPR	136	B-C-IS-ITND
37	IS-WSI-F	137	B-C-IS-ITT
38	B-D-IS-WSI	138	C-IS-ITD
39	B-E36-IS-WSI	139	C-IS-ITND
40	B-C-D-IS-WSI	140	C-IS-ITT
41	B-IS-WSI	141	C-ORD-IS-ITD
42	B-TO-E36-IS-WSI	142	C-ORD-IS-ITND
43	C-IS-WSI	143	C-ORD-IS-ITT
44	D35-E36-IS-WSI	144	MIG-CAG-IS-ITD
45	D-IS-WSI	145	MIG-CAG-IS-ITND
46	E36-IS-WSI	146	MIG-CAG-IS-ITT
47	MIG-CAG-IS-WSI	147	MIG-COG-IS-ITD
48	MIG-COG-IS-WSI	148	MIG-COG-IS-ITND
49	MIG-DCOG-IS-WSI	149	MIG-COG-IS-ITT
50	MIG-ING-IS-WSI	150	MIG-DCOG-IS-ITD
51	MIG-NDCOG-IS-WSI	151	MIG-DCOG-IS-ITND
52	MIG-NRG-IS-WSI	152	MIG-DCOG-IS-ITT
53	BS-CCI-BAL	153	MIG-ING-IS-ITD

Labels and Groups of Variables used

#	Label	#	Label
54	BS-CEME-BAL	154	MIG-ING-IS-ITND
55	BS-COB-BAL	155	MIG-ING-IS-ITT
56	BS-CPE-BAL	156	MIG-NDCOG-IS-ITD
57	BS-CTA-BAL	157	MIG-NDCOG-IS-ITT
58	BS-BCI	158	IS-PEI-F-CC11-X-CC113
59	BS-CSMCI	159	IS-IPI-F-CC11-X-CC113
60	BS-FS-LY	160	IS-HWI-F
61	BS-FS-NY	161	C-ORD-IS-IO
62	BS-GES-LY	162	C-ORD-X-C30-IS-IO
63	BS-GES-NY	163	IS-CAR
64	BS-MP-NY	164	FOOD-IS-DIT
65	BS-MP-PR	165	IS-DIT
66	BS-PT-LY	166	NFOOD-IS-DIT
67	BS-PT-NY	167	NFOOD-X-G473-IS-DIT
68	BS-SFSH	168	X-G473-IS-DIT
69	BS-SV-NY	169	M1
70	BS-SV-PR	170	M2
71	BS-UE-NY	171	M3
72	BS-ICI	172	3MI-RT
73	BS-IEME	173	LTGBY-RT
74	BS-IEOB	174	EXA-RT-USD
75	BS-IOB	175	EXA-RT-JPY
76	BS-IPE	176	EXA-RT-GBP
77	BS-IPT	177	BDSHRPRCF
78	BS-ISFP	178	DJES50I
79	BS-ISPE	179	EMECB2Y
80	BS-RAS	180	EMECB3Y
81	BS-RCI	181	EMECB5Y
82	BS-REBS	182	EMECB7Y
83	BS-REM	183	EMGBOND
84	BS-ROP	184	FIBOR1Y
85	BS-RPBS	185	FIBOR3M
86	BS-SABC	186	FIBOR6M
87	BS-SAEM	187	BDWU1032R
88	BS-SARM	188	BDWU0022R
89	BS-SCI	189	BDEBDBSIA
90	BS-SERM	190	BDECBXDGA
91	BS-CSMCI-B	191	BDECBXDMA
92	BS-ESI-I	192	BDECBXNGA
93	BS-ICI-BAL	193	BDECBXNOA
94	BS-RCI-BAL	194	BDECBXOLA
95	BS-SCI-BAL	195	BDECBXLIA
96	RT-LM-UN-T-GT25		
97	RT-LM-UN-T-LE25		
98	RT-LM-UN-T-TOT		
99	1000-PERS-LM-UN-T-GT25		
100	1000-PERS-LM-UN-T-LE25		

16.C Large dataset of predictors: Transformations

Transformations

#	Label	#	Label	#	Label	#	Label
1	First Diff., Logs	51	First Diff., Logs	101	No Change	151	First Diff., Logs
2	First Diff., Logs	52	First Diff., Logs	102	First Diff., Logs	152	First Diff., Logs
3	First Diff., Logs	53	No Change	103	First Diff., Logs	153	First Diff., Logs
4	First Diff., Logs	54	No Change	104	First Diff., Logs	154	First Diff., Logs
5	First Diff., Logs	55	No Change	105	First Diff., Logs	155	First Diff., Logs
6	First Diff., Logs	56	No Change	106	First Diff., Logs	156	First Diff., Logs
7	First Diff., Logs	57	No Change	107	First Diff., Logs	157	First Diff., Logs
8	First Diff., Logs	58	No Change	108	First Diff., Logs	158	First Diff., Logs
9	First Diff., Logs	59	No Change	109	First Diff., Logs	159	First Diff., Logs
10	First Diff., Logs	60	No Change	110	First Diff., Logs	160	First Diff., Logs
11	First Diff., Logs	61	No Change	111	First Diff., Logs	161	First Diff., Logs
12	First Diff., Logs	62	No Change	112	First Diff., Logs	162	First Diff., Logs
13	First Diff., Logs	63	No Change	113	First Diff., Logs	163	First Diff., Logs
14	First Diff., Logs	64	No Change	114	First Diff., Logs	164	First Diff., Logs
15	First Diff., Logs	65	No Change	115	First Diff., Logs	165	First Diff., Logs
16	First Diff., Logs	66	No Change	116	First Diff., Logs	166	First Diff., Logs
17	First Diff., Logs	67	No Change	117	First Diff., Logs	167	First Diff., Logs
18	First Diff., Logs	68	No Change	118	First Diff., Logs	168	First Diff., Logs
19	First Diff., Logs	69	No Change	119	First Diff., Logs	169	First Diff., Logs
20	First Diff., Logs	70	No Change	120	First Diff., Logs	170	First Diff., Logs
21	First Diff., Logs	71	No Change	121	First Diff., Logs	171	First Diff., Logs
22	First Diff., Logs	72	No Change	122	First Diff., Logs	172	First Diff.
23	First Diff., Logs	73	No Change	123	First Diff., Logs	173	First Diff.
24	First Diff., Logs	74	No Change	124	First Diff., Logs	174	First Diff.
25	First Diff., Logs	75	No Change	125	First Diff., Logs	175	First Diff.
26	First Diff., Logs	76	No Change	126	First Diff., Logs	176	First Diff.
27	First Diff., Logs	77	No Change	127	First Diff., Logs	177	First Diff., Logs
28	First Diff., Logs	78	No Change	128	First Diff., Logs	178	First Diff., Logs
29	First Diff., Logs	79	No Change	129	First Diff., Logs	179	First Diff.
30	First Diff., Logs	80	No Change	130	First Diff., Logs	180	First Diff.
31	First Diff., Logs	81	No Change	131	First Diff., Logs	181	First Diff.
32	First Diff., Logs	82	No Change	132	First Diff., Logs	182	First Diff.
33	First Diff., Logs	83	No Change	133	First Diff., Logs	183	First Diff.
34	First Diff., Logs	84	No Change	134	First Diff., Logs	184	First Diff.
35	First Diff., Logs	85	No Change	135	First Diff., Logs	185	First Diff.
36	First Diff., Logs	86	No Change	136	First Diff., Logs	186	First Diff.
37	First Diff., Logs	87	No Change	137	First Diff., Logs	187	First Diff.
38	First Diff., Logs	88	No Change	138	First Diff., Logs	188	First Diff.
39	First Diff., Logs	89	No Change	139	First Diff., Logs	189	First Diff., Logs
40	First Diff., Logs	90	No Change	140	First Diff., Logs	190	First Diff., Logs
41	First Diff., Logs	91	No Change	141	First Diff., Logs	191	First Diff., Logs
42	First Diff., Logs	92	No Change	142	First Diff., Logs	192	First Diff., Logs
43	First Diff., Logs	93	No Change	143	First Diff., Logs	193	First Diff., Logs
44	First Diff., Logs	94	No Change	144	First Diff., Logs	194	First Diff., Logs
45	First Diff., Logs	95	No Change	145	First Diff., Logs	195	First Diff., Logs
46	First Diff., Logs	96	First Diff.	146	First Diff., Logs		
47	First Diff., Logs	97	First Diff.	147	First Diff., Logs		
48	First Diff., Logs	98	First Diff.	148	First Diff., Logs		
49	First Diff., Logs	99	No Change	149	First Diff., Logs		
50	First Diff., Logs	100	No Change	150	First Diff., Logs		

Bibliography

Alvarez R., Camacho M., Perez-Quiros G. (2012), Finite sample performance of small versus large scale dynamic factor models, Discussion Paper 8867, Centre for Economic Policy Research.

Aruoba B., Diebold F., Scotti C. (2009), Real-time measurement of business conditions, Journal of Business and Economic Statistics, 27, 417-427.

Aruoba B., Diebold F. (2010), Real-time macroeconomic monitoring: Real activity, inflation and interactions, American Economic Review, 100, 20-24.

Ng S., Boivin J. (2006), Are more data always better for factor analysis? Journal of Econometrics, 132, 169-194.

Camacho M., Perez-Quiros G. (2010), Introducion the Euro-STING: Short Term INdicator of Euro Area Growth, Journal of Applied Econometrics, 25, 663-694.

Camba-Mendez G., Kapetanios G., Smith R.J., Weale M. R. (2001), An automatic leading indicator of economic activity: forecasting GDP growth for European countries, Econometrics Journal, 4, S56-S90.

De Mol C., Giannone D., Reichlin L. (2006), Forecasting with a large number of predictors: is bayesian regression a valid alternative to principal components?, Discussion Paper 5829, Centre for Economic Policy Research.

Fernandez C., Ley E., Steel M.F.J. (2001), Benchmark priors for bayesian model averaging, Journal of Econometrics, 100, 381-427.

Forni M., Hallin M., Reichlin L. (2005), The generalized dynamic factor model: one-sided estimation and forecasting, Journal of the American Statistical Association, 100, 830-840.

Foroni C., Marcellino M. (2011), A comparison of mixed frequency approaches for modelling euro area macroeconomic variables, Mimeo.

Foroni C., Marcellino M., Schumacher C. (2011), MIDAS and unrestricted MIDAS models, Mimeo.

Goffe W.L., Ferrier G.D., Rogers J. (1994), Global optimisation of statistical functions with simulated annealing, Journal of Econometrics, 60-1, 65-99.

Hajek B. (1998), Cooling schedules for optimal annealing, Mathematics of Operations Research, 13-2, 311–331.

Hannan E.J., Quinn B.G. (1979), The determination of the order of an autoregression, Journal of the Royal Statistical Society (Series B), 41, 190-195.

Hartl H.R.F., Belew R.K. (1990), A global convergence proof for a class of genetic algorithms, Technical Report,Technical Universiy of Vienna.

Harvey D. (1993), Time Series Models, Hemel Hempstead: Harvester and Wheatsheaf.

Helland I.S. (1998), On the structure of partial least squares regression, Communications in Statistics - Simulation and Computation, 17, 581-607.

Helland I.S. (1990), Partial least squares regression and statistical models, Journal of Statistics, 17, 91-114.

Hoover K.D., Perez S.J. (1999), Data mining reconsidered: Encompassing and the general-to-specific approach to specification set, Econometrics Journal, 2, 167-191.

Kapetanios G. (2006), Variable selection in regression models using non-standard optimisation of information criteria, Computational Statistics and Data Analysis, 52-1, 4-15.

Kapetanios G., Marcellino M., Papailias F. (2014a), Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods, quantf research Working Paper Series: WP04/2014, available onine at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2444418.

Kapetanios G., Marcellino M., Papailias F. (2014b), Forecasting EU economic activity using financial conditions indexes, quantf research Working Paper Series: WP06/2014, available onine at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2444416.

Kapetanios G., Marcellino M., Papailias F. (2014c), Business cycles dating for EU economies: an empirical search for the optimal settings, quantf research Working Paper Series: WP05/2014, available onine at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2444415.

Kapetanios G., Marcellino M., Papailias F. (2014d), Forecasting EU economic activity using summary indicators, quantf research Working Paper Series: WP07/2014, available onine at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2444422.

Morinaka Y., Yoshikawa M., Amagasa T. (2001), The L-Index: An indexing structure for efficient subsequence matching in time sequence databases, Proceedings of the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining.

Schwarz G. (1978), Estimating the dimension of the model, Annals of Statistics, 6, 461-464.

Sin C.Y., White H. (1996), Information criteria for selecting possibly misspecifed parametric models, Journal of Econometrics, 71-(1-2), 207-225.

Stock J., Watson M. (1989), New indexes of coincident and leading economic indicators, NBER Macroeconomic Annual, 351-394.

Stock J., Watson M. (1991), A probability model of the coincident economic indicators, In Leading Economic Indicators: New Approaches and Forecasting Records, edited by K. Lahiri and G. Moore, Cambridge University Press.

Stock J., Watson M. (2002a), Forecasting using principal components from a large number of predictors, Journal of the American Statistical Association, 297, 1167-1179.

Stock J., Watson M. (2002b), Macroeconomic forecasting using diffusion indexes, Journal of Business & Economics Statistics, 20, 147-162.

Wold H. (1982), Soft modeling, The basic design and some extensions, In: Joreskog, K. G., Wold, H., Systems under indirect observation, Vol. 2, North-Holland, Amsterdam.

17

Large-Dimensional Dynamic Factor Models in Real-Time

Author

Matteo Luciani

Board of Governors of the Federal Reserve System

matteo.luciani@frb.gov

Table of Contents

17.1 Introduction

In the last fifteen years Large-Dimensional Dynamic Factor models have become increasingly popular in economic literature. Due to the strong co-movement among macroeconomic time series, these models offer a parsimonious and realistic representation of the data, and this is why they have proven successful in forecasting Stock and Watson (2002a), Stock and Watson (2002b), Forni et al. (2005), Giannone et al. (2008); Luciani (2014), in construction of both business cycle indicators Stock and Watson (1999), Altissimo et al. (2001, 2010) and core inflation indexes Cristadoro et al. (2005), in structural analysis Giannone et al. (2005), Stock and Watson (2005), Forni et al. (2009), Forni and Gambetti (2010), Barigozzi et al. (2014), Luciani (2015), as well as in the analysis of financial markets Corielli and Marcellino (2006), Ludvigson and Ng (2007), Ludvigson and Ng (2009), Hallin et al. (2011), Luciani and Veredas (2015).

Factor models have been a common tool in many scientific fields but until recent years achieved moderate success in economic analysis more likely because they were estimable only on small databases. However, once the literature understood how to estimate them under general assumptions, such as those needed for a large economic database, they have become a standard tool in macroeconomic literature.

Until the late nineties, estimation of Factor models was possible only under the assumption that the co-movement in the data is the result of macroeconomic shocks (as for example monetary policy or technology shocks), an assumption that clearly does not apply to large economic databases where sectorial or regional shocks are likely to affect groups of variables. Then, in the late nineties the seminal works of Forni et al. (2000) and of Stock and Watson (2002a) have demonstrated that, as both the sample size and the number of variables diverge to infinity, Factor models can be consistently estimated with the method of static/dynamic principal components, even under the assumption that the correlation in the data is due not only to macroeconomic shocks, but also to other *non-pervasive* shocks. Furthermore, lately Doz et al. (2011, 2012) have shown that the techniques used for estimating small models, such as the Kalman Filter or the Expectation-Maximization algorithm, are, indeed, relevant also for the case of Large Dynamic Factor models.

The key success of Factor models in economics has been the fact that they can be estimated on large databases without suffering the curse of dimensionality. There are two main reasons why economists desire a model that is estimable on large databases: the first is a consequence of the common point of view among professional forecasters according to which a model with a large number of predictors may improve the forecast of macroeconomic indicators. The second reason, instead, comes from the observation that professional literature, such as economic reports from central banks or other economic institutions, typically provides analyses of the behavior of a large number of series. This suggests that policy makers consider all of them to contain significant information about the state of the economy, thus deserving to be included in the forecasting model.

The goal of this chapter is to review the literature on Large-Dimensional Dynamic Factor Models for real-time applications, and in particular for constructing business cycle indicators and for predicting economic activity. For a detailed analysis of Small-Dimensional Dynamic Factor Models we refer the reader to the review by Filippo Moauro (Chapter 11 of this handbook).

In addition to the standard challenges of factor analysis, the use of these models in real-time and on large databases presents some extra problems mainly due to the fact that key economic statistics are released with a long delay, and that they are also subsequently revised. Moreover, different types of economic data are released in a non-synchronous manner, with different degrees of delay, and at different frequencies. The goal of this Chapter is to review the techniques that the literature has suggested to overcome these problems and to present some empirical applications.

This Chapter is organized as follows: in Section 17.2, we start by presenting the Factor model and in particular by focusing on the representation of the economy that this model is imposing, and on the feature of the data that it is mostly able to capture. Then, we discuss estimation, first by explaining what are the problems induced by the use of large databases and how these problems can be solved by non-parametric methods, and then

by discussing maximum likelihood estimation. Finally, we conclude the discussion on the model by discussing the challenges of real-time applications.

In Section 17.3 we discuss the different variations of the model presented in 17.2 which have been used for predicting economic activity. In this Section, we also present some empirical applications, and in particular in Section 17.3.1 we present results from the literature on *nowcasting* economic activity. Then, in Section 17.4 we present two examples of business cycles indicators constructed with Large-Dimensional Factor models.

In Section 17.5 we discuss issues related with large-dimensional databases, and in particular we try to understand whether large databases are really useful. Surprisingly so, the conclusion is that the common viewpoint according to which more variables imply better forecasts, is not necessarily true. The intuitive explanation for this result is that since there is a lot of co-movement in macroeconomic data, often by adding variables one end up adding redundant information. In other words, there is so much co-movement in macroeconomic data that if one can select correctly 15-20 variables, then the informational content left in the other available variables is marginal. Therefore, the conclusion of this Section is that unless one needs to comment on many data releases, there is no need of a large database when forecasting with Dynamic Factor models, of course as long as the variables on which the model is estimated are appropriately selected. Finally, Section 17.6 concludes.

This chapter contains also a technical annex addressing all the technicalities that we have backed out from the main text for readability purposes, such as, for example, Generalized Principal Components and the Expectation-Maximization algorithm, as well as two extensions of the model presented in Section 17.2, the first in a Bayesian framework, and the second in a non-stationary set-up.

17.2 Dynamic Factor Models

17.2.1 Representation

Factor models are based on the idea that macroeconomic fluctuations are the result of a small number of macroeconomic/structural shocks ($\mathbf{u}_t$), which affect all the variables, and of a large number of sectorial/regional shocks ($\mathbf{e}_t$) that affect one or a few variables. Therefore, each variable in the dataset (x_{it}) can be decomposed into the sum of a common (χ_{it}) and an idiosyncratic (ξ_{it}) component. Formally, let $\mathbf{x}_t$ be a vector of n stationary variables observed at month t, the Dynamic Factor model is defined as follows:

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t \tag{17.1}$$

$$\boldsymbol{\chi}_t = \mathbf{C}(L)\mathbf{u}_t \tag{17.2}$$

$$\boldsymbol{\xi}_t = \mathbf{D}(L)\mathbf{e}_t \tag{17.3}$$

where $\mathbf{u}_t$ and $\mathbf{e}_t$ are white noise processes, respectively, of dimension $q \times 1$ and $n \times 1$, while $\mathbf{C}(L) = \sum_{j=0}^{\infty} C_j L^j$ and $\mathbf{D}(L) = \sum_{j=0}^{\infty} D_j L^j$ are $n \times q$ and $n \times n$ square summable polynomials in the lag operator.

The common shocks and the idiosyncratic shocks are two independent sources of fluctuations, and hence they are assumed to be uncorrelated at all leads and lags: $\mathbb{E}(u_{it}e_{js}) = 0 \ \forall i, j, s, t$. What differentiate common and idiosyncratic shocks is that the former are *pervasive* (meaning that they affect all the variables in the database), while the latter are *non-pervasive* (meaning that they affect only a limited number of variables). This distinction will be formalized in Annex 17.A.1, in the meantime what matters is that the bulk of the cross-correlation in the data is captured by the common component, whereas the idiosyncratic components are just weakly cross-sectionally correlated (approximate factor structure).

Model (17.1)-(17.3) is known in the literature as the *Generalized* or the *Unrestricted* Dynamic Factor model and it is studied in Forni et al. (2000), Forni et al. (2004), Forni et al. (2014) and Forni and Lippi (2001), Forni and Lippi (2010). For a formal treatment of this model, including estimation, we refer the reader to these

papers. We presented this model here just for exposition purposes since it is the model that better explicates the intuition behind Factor models. In fact, since estimation of (17.1)-(17.3) is problematic, this model is not the Dynamic Factor model used in the forecasting literature. Rather, the model used in the literature is a restricted version of model (17.1)-(17.3) where the responses of the variables to the common shocks, $\mathbf{C}(L)$, are restricted to be proportional across the xs. That is, it is assumed that, although the combination of shocks that generate fluctuations is different between different cycles, the relative movements of the variables $\mathbf{x}_t$ in response to these shocks is approximately the same.

Formally, this idea is made explicit by assuming that $\mathbf{C}(L) = \mathbf{\Lambda B}(L)$, where $\mathbf{\Lambda}$ is $n \times r$, $\mathbf{B}(L) = \sum_{j=0}^{\infty} B_j L^j$ is an $r \times q$ square summable polynomial in the lag operator, and $q \leq r$. With this restriction model (17.1)-(17.3) can be rewritten as

$$\mathbf{x}_t = \mathbf{\chi}_t + \xi_t \tag{17.4}$$

$$\mathbf{\chi}_t = \mathbf{\Lambda F}_t \tag{17.5}$$

$$\mathbf{F}_t = \mathbf{B}(L)\mathbf{u}_t \tag{17.6}$$

$$\mathbf{\xi}_t = \mathbf{D}(L)\mathbf{e}_t \tag{17.7}$$

where $\mathbf{F}_t$ is an $r \times 1$ vector containing the common factors, which are meant to capture the comovement in the data, *i.e.* the business cycle. This model is known in the literature as the *Restricted* Dynamic Factor model or simply as the Dynamic Factor model, and it is studied in Stock and Watson (2005), Forni et al. (2009), and Doz et al. (2011, 2012) among others. In addition to the discussion in the rest of this chapter, we refer the interested reader to these references.

17.2.2 Estimation

Issues in Estimating Large-Dimensional Dynamic Factor Models

Pioneered by the work of Charles E. Spearman, Factor models are a statistical technique applied in many fields since the beginning of the twentieth century. Until recently, however, Factor models in economics were used only on small databases Geweke (1977), Sargent and Sims (1977), Engle and Watson (1981), Stock and Watson (1989).[1] This is because, estimation of Factor models was possible only under the hypothesis that the idiosyncratic components are cross-sectionally uncorrelated (exact factor structure), that is under the assumption that all the correlation in the dataset is due to the common shocks. This assumption is unrealistic on large databases where sectorial or regional shocks might affect groups of variables, but it may be plausible on small databases of aggregate macroeconomic variables.

When the factor structure is exact, model (17.4)-(17.7) can be estimated with likelihood methods. Depending on the assumptions on the law of motion of the factors (17.6) and on the law of motion of the idiosyncratic components (17.7), maximum likelihood estimation can be obtained by closed formulas Lawley and Maxwell (1971), or by using the Kalman Filter and then maximizing the likelihood either with numerical methods, or with the EM algorithm Watson and Engle (1983).

When the factor structure is approximate, estimation is trickier. Indeed, the common point of view was that in this set-up estimation by likelihood methods is no longer feasible because there are too many parameters to be estimated. This point of view led researchers to develop non-parametric estimation techniques to estimate (17.4)-(17.7).[2] However, recently, Doz et al. (2012) proved that this common point of view was rather a

---

[1]See Chapter 9 for an extensive treatment of Small-Dimensional Dynamic Factor Models.

[2]The effort of many scholars has produced different non-parametric estimators. In this chapter we will discuss only Static Principal Components Stock and Watson (2002a), Bai and Ng (2002), Bai (2003) and Generalized Principal Components Forni et al. (2005), Choi (2012), while we refer the reader to Forni et al. (2000), Forni et al. (2004) for Dynamic Principal Components estimation, and to Breitung and Tenhofen (2011), 2011 for GLS estimation.

common misconception since estimation based on likelihood methods of (17.4)-(17.7) under cross-correlated idiosyncratic components is, indeed, possible.

Estimation by Principal Components

The main intuition of the approximate factor literature is that as the number of variables increases to infinity, the common component survives to aggregation whereas the idiosyncratic component vanishes. Let us better explain this intuition with a stylized example.

Suppose that the true model is a factor model with one factor and with all the loadings equal to one:

$$x_{it} = f_t + \xi_{it},$$

where $\mathbb{E}(f_t) = 0$ and $\mathbb{E}(\xi_{it}) = 0$. Then, if we take the cross-sectional average of the data, that is $\bar{x}_t = \frac{1}{n}\sum_{i=1}^{n} x_{it}$, we obtain a consistent estimator of f_t. Indeed, we can rewrite $\bar{x}_t$ as:

$$\bar{x}_t = \frac{1}{n}\sum_{i=1}^{n} f_t + \frac{1}{n}\sum_{i=1}^{n} \xi_{it}$$

$$= f_t + \frac{1}{n}\sum_{i=1}^{n} \xi_{it}$$

and then, for n going to infinity, the term $\frac{1}{n}\sum_{i=1}^{n} \xi_{it}$ goes to zero for the weak law of large numbers. Hence, if $\lim_{n\to\infty} \bar{x}_t = f_t$, that is $\bar{x}_t$ is a consistent estimate of f_t.[3]

Obviously, this was an extremely stylized example that we used just for illustration, and in practice factors are not estimated by simple cross-sectional averages. However, despite being stylized, this example it is not too far away from what the literature has suggested to do since factors are estimated by *weighted* cross-sectional averages. The question is: what kind of weights should we use? As we explain in Annex 17.A.1, the weights should be such that they maximize the variance of the cross-sectional averages.

Formally, let $\mathbf{w}$ be a vector of weights with $\frac{\mathbf{w}'\mathbf{w}}{n} = \mathbf{I}_r$, then $\mathbf{w}$ should be such that the variance of the weighted cross-sectional average $\frac{1}{n}\mathbf{w}'\mathbf{x}_t$. The solution to this problem is equivalent to finding the first r principal components of $\mathbf{x}_t$, that is:

$$\widehat{\mathbf{\Lambda}} = \mathcal{V}_r \mathcal{D}_r^{1/2} \tag{17.8}$$

$$\widehat{\mathbf{F}}_t = \frac{1}{n}\widehat{\mathbf{\Lambda}}'\mathbf{x}_t, \tag{17.9}$$

where $\mathcal{D}_r$ is an $r \times r$ diagonal matrix containing the eigenvalues, and $\mathcal{V}_r$ is the $n \times r$ matrix containing the associated eigenvectors, of the covariance matrix of $\mathbf{x}_t$.

Estimation of Approximate Factor models with (static) principal components is studied in Stock and Watson (2002a), Bai and Ng (2002), Bai (2003), and Forni et al. (2009). We refer the interested reader to these references for more details. For the purpose of this Chapter what matters is that this estimator is consistent for both n and T diverging to infinity.

---

[3] Actually for consistency it is also necessary that asymptotically the variance of $\bar{x}_t$ is equal to that of f_t, that is asymptotically $\bar{x}_t$ captures the same variation of f_t. Let us assume that $var(f_t) = 1$, $var(\xi_{it}) = \psi_i^2$, and $cov(\xi_{it}, \xi_{ij}) = \psi_{ij}$, then the variance of $\bar{x}_t$ is equal to $var(\bar{x}_t) = 1 + \frac{1}{n^2}\sum_{i=1}^{n} \psi_i^2 + \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n-1} \psi_{ij}$. Now, $\lim_{n\to\infty} var(\bar{x}_t) = 1$ if and only if $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \psi_i^2 + \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n-1} \psi_{ij} = 0$, which is the case when the idiosyncratic shocks are non-pervasive and the hence the idiosyncratic components are just weakly cross-sectionally correlated.

Estimating the Dynamics of the Model

Once the factors, the loadings, and the idiosyncratic component are estimated it is still necessary to estimate the filters $\mathbf{B}(L)$ and $\mathbf{D}(L)$ in (17.6) and (17.7), that is, it is necessary to estimate a law of motion for the factors and a low of motion for the idiosyncratic components.

With respect to the common factors, it is common practice to assume that the vector of $\mathbf{F}_t$ evolves over time according to a VAR model, $\mathbf{A}(L)\mathbf{F}_t = \mathbf{B}\mathbf{u}_t$ (where $\mathbf{B}$ is $r \times q$), and to estimate it with OLS. Note that this is a mild assumption since VAR models are very general models that are able to capture a wide range of dynamics. Moreover, this is an assumption only when $q = r$, because when (i) $q < r$ and (ii) $\mathbf{F}_t$ has rational spectral density, the VAR representation *generically* exists Anderson and Deistler (2008a), Anderson and Deistler (2008b).

With respect to the idiosyncratic component, when forecasting it is common practice to assume either that each ξ_{it} evolves according to an autoregressive process, or that $\xi_{it} \sim iid(0, \psi_i)$. At a first glance this modeling choice may look inconsistent, since we estimate the model under the hypothesis of cross-correlated idiosyncratic components, but then in forecasting we use it as if the idiosyncratic components were mutually orthogonal. However, results in Luciani (2014) show that accounting for cross-correlation due to non-pervasive shocks in the idiosyncratic component rarely boosts the forecasting accuracy.

Estimation by Likelihood Methods

Doz et al. (2012) show that the same techniques that are used for estimation of Small Exact Dynamic Factor models, can be used also for Large Approximate Dynamic Factor models. Namely, Doz et al. (2012) prove that model (17.4)-(17.7) can be consistently estimated by Quasi Maximum Likelihood. The intuition of Doz et al. (2012) is that the exact factor model can be treated as a miss-specified approximating model, that is the model can be estimated with maximum likelihood as if the idiosyncratic components were orthogonal. Indeed, as shown by Doz et al. (2012) when n and T diverge to infinity the miss-specification error vanishes.

For example, Doz et al. (2012) show that under the assumption that $\mathbf{A}(L)\mathbf{F}_t = \mathbf{u}_t$, with $\mathbf{u}_t \sim \mathcal{N}(0, \mathbf{I}_r)$, and $\xi_t \sim \mathcal{N}(0, \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a diagonal matrix, the model can be consistently estimated by the EM algorithm presented in Annex 17.A.3.

As another example, Doz et al. (2012) show that even under an approximating model which assume $\mathbf{F}_t \sim \mathcal{N}(0, \mathbf{I}_r)$ and $\xi_t \sim \mathcal{N}(0, \psi^2 \mathbf{I}_n)$ (that is a highly miss-specified model), the factors and the loadings can be consistently estimated as

$$\widehat{\mathbf{\Lambda}} = \mathcal{V}_r(\mathcal{D}_r - \widehat{\psi}^2 \mathbf{I}_r)^{1/2} \tag{17.10}$$

$$\widehat{\mathbf{F}}_t = \mathbb{E}(\mathbf{F}_t | \mathbf{x}_t) = \mathcal{D}^{-1} \widehat{\mathbf{\Lambda}}' \mathbf{x}_t \tag{17.11}$$

where $\widehat{\psi}^2 = \frac{1}{n} \mathrm{Trace}(\mathbf{\Sigma}^x - \widehat{\mathbf{\Lambda}} \widehat{\mathbf{\Lambda}}').$[4]

17.2.3 Large-Dimensional Factor Models in Real-Time

Factor models have been used in real-time both for constructing business cycle indicators, and for predicting economic activity. When constructing business cycle indicators, the goal is to capture the fluctuations in the data due to "fundamentals", i.e. to the common shocks $\mathbf{u}_t$. In other words, the goal is to extract the main "signal" in the data while backing out the "noise". When predicting economic activity, instead, the goal is simply to find the best prediction of a given variable of interest such as, for example, GDP, Industrial Production, or the Consumer Price Index.

---

[4]Note that since the factors and the loadings estimated as in (17.10) and (17.11) are proportional to those estimated as in (17.8) and (17.9), then the principal component estimator can be reinterpreted as a Quasi Maximum Likelihood estimator.

Roughly speaking, Factor models are used in real-time as follows: (i) supposes that we are at date v and that at date v it is available a given vintage of data Ω_v; (ii) further supposes that on the basis of this vintage of data we constructed either our indicator, or our prediction. (iii) now, supposes that at day $v+1$ a new data is released (*eg.* PMI), (iv) then on the basis of the new vintage of data Ω_{v+1} we can update our estimate of the factors, and hence we can update our indicator/prediction.

During the past years different versions of model (17.4)-(17.7) have been used for real-time applications. In addition to the standard challenges of factor analysis, the use of model (17.4)-(17.7) in real-time presents two main problems:

1. how to manage *missing values*, in particular at the end of the sample due to the delay in macroeconomic data releases, and

2. how to handle *mixed frequencies*, which most of the times reduces to how to bridge monthly and quarterly indicators.

In the next two Sections we will present the different versions of model (17.4)-(17.7) used for constructing business cycle indicators and for predicting economic activity by separating them depending on the task they are used for. In particular, we will concentrate on how the *missing values* and the *mixed frequencies* problems were addressed.

17.3 Predicting Economic Activity

There exists a large literature on predicting economic activity with Factor models and this literature has shown that these models outperform standard univariate benchmark predictions. The goal of this Section is to present the different approaches used for predicting with Dynamic Factor models in real-time.[5]

The first approach we discuss is the Diffusion Index model suggested by Stock and Watson (2002a), Stock and Watson (2002b). Let $\mathbf{x}_t$ be the vector of potential predictors, and let y_t be the target variable to be forecasted, Stock and Watson (2002a), Stock and Watson (2002b) assume that $(\mathbf{x}_t, y_t)$ have an approximate factor representation, that is $\mathbf{x}_t$ and y_t that are driven by the same common factors, namely:

$$y_{t+h} = c + \boldsymbol{\beta}(L)\mathbf{F}_t + \alpha(L)y_t \tag{17.1}$$

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t, . \tag{17.2}$$

Stock and Watson (2002a), Stock and Watson (2002b) suggest to estimate the common factors in (17.2) with principal components as explained in Section 17.2.2, and then to use the estimated factors to predict y_T, which practically means estimating (17.1) with OLS, where $\mathbf{F}_t$ is replaced by $\widehat{\mathbf{F}}_t$. Moreover, Bai and Ng (2006) proved that, if $\sqrt{T}/N \to 0$, the OLS estimate of $\beta(L)$ and $\alpha(L)$ are $\sqrt{T}$ consistent even though the true factors are replaced by principal components estimates.

Stock and Watson (2002a), Stock and Watson (2002b) use model (17.1)-(17.2) to perform a pseudo real-time forecasting exercise. Their goal is to predict two monthly indicators (Industrial Production and Inflation), and to do this they use only monthly indicators. In other words, Stock and Watson (2002a), Stock and Watson (2002b) have neither missing values, nor mixed frequencies. However, Stock and Watson (2002b) propose an EM algorithm (described in Annex 17.A.3) to account for missing values as well as for mixed frequency in $\mathbf{x}_t$.

---

[5]There exists a vast literature, which we will not review in this Chapter, that used Dynamic Factor models for forecasting in *pseudo* real-time. A (non-exhaustive) list of papers that used this approach is: Stock and Watson (2002a), Stock and Watson (2002b), Forni et al. (2003), Forni et al. (2005), Marcellino et al. (2003), Boivin and Ng (2005), Boivin and Ng (2006), Artis et al. (2005), Schumacher (2007), D'Agostino and Giannone (2012), and Luciani (2014).

Model (17.1)-(17.2) is used in real-time by Bernanke and Boivin (2003) to forecast CPI inflation, industrial production, and the unemployment rate. In addition to accounting for missing values at the end of the sample, Bernanke and Boivin (2003) include both quarterly and monthly variables in $\mathbf{x}_t$, and they estimate the model with the EM-algorithm of Stock and Watson (2002b). Table 17.1 reports results from Bernanke and Boivin (2003), which point out that, for inflation, model (17.1)-(17.2) does similarly to an AR model, while for Industrial Production, and the Unemployment rate it does, respectively, 10-15% and 15-20% better than the AR model.

Table 17.1: Relative Mean Squared Forecasting Error of model (17.1)-(17.2)

h	CPI	IP	UR
6	0.98	0.84	0.86
12	0.96	0.92	0.8

This Table shows relative mean squared forecasting error of model (17.1)-(17.2) relative to an AR model at the 6-month horizon and at the 12-month horizon. The target variables are CPI inflation, industrial production, and the unemployment rate, and the model is estimated on a real-time database of 78 variables including both quarterly and monthly indicators.
Source: Bernanke and Boivin (2003).

A slightly different model is used by Giannone et al. (2008), namely

$$y_t^Q = c + \boldsymbol{\beta}\mathbf{F}_t \tag{17.3}$$

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{17.4}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{B}\mathbf{u}_t \tag{17.5}$$

where y_t^Q is quarter-on-quarter GDP growth observed quarterly (i.e. only at $t = 3, 6, 9, \ldots T$), $\mathbf{x}_t$ is a vector of monthly indicators transformed so that they correspond to a quarterly quantity when observed at the end of the quarter,[6] and $\mathbf{B}$ is of dimension $r \times q$.

The main difference with the model of Stock and Watson (2002a), Stock and Watson (2002b) is that Giannone et al. (2008) estimate (17.4)-(17.5) with the two-step estimator studied in Doz et al. (2011) consisting in using Principal components and the Kalman Filter. Similarly to Stock and Watson (2002a), Stock and Watson (2002b), Giannone et al. (2008) estimate (17.3) by OLS, i.e. by projecting quarter GDP onto the space spanned by the estimated quarterly factors. Missing values at the end of the quarter are handled with the Kalman Filter (see Annex 17.A.3), which is also used for computing expectations of the factors.[7]

A third approach, studied in Bańbura and Modugno (2014), consists in modeling quarterly and monthly variables jointly. Let $\mathbf{x}_t$ be a vector of both quarterly and monthly variables, then the model of Bańbura and Modugno (2014) is as follows:

$$\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{17.6}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{B}\mathbf{u}_t \tag{17.7}$$

$$\xi_{it} = \alpha_i\xi_{it-1} + e_{it} \qquad i = 1, \ldots, n. \tag{17.8}$$

Suppose, without loss of generality, that the target variable is x_{1t} and that the goal is to forecast one-step ahead. Then, the optimal forecast is $x_{1t+1} = \boldsymbol{\Lambda}_1\mathbf{A}\mathbf{F}_t + \alpha_1\xi_{1t}$, which requires estimation of all the parameters in (17.6)-(17.8).

Bańbura and Modugno (2014) propose an EM algorithm (presented in Annex 17.A.3) to estimate (17.6)-(17.8) which account for both mixed frequency and missing values. An alternative way of estimating (17.6)-(17.8) is to

---

[6]Let X_{it} be a generic monthly variable in levels, then this variable can be converted in quarterly observations as: $X_{it}^Q = X_{it} + X_{it-2} + X_{it-2}$. Then, the transformation used by Giannone et al. (2008) is: $x_{it} = X_{it}^Q - X_{it-3}^Q$.

[7]For example, suppose we are in January, so that T is January, and that we need to predict Q1 GDP growth (y_{T+2} in our notation) by using (17.3). In order to do that we need a forecast of $\mathbf{F}_{T+2}$, and this forecast is obtained recursively by the Kalman Filter.

follow Schumacher and Breitung (2008) who use the EM algorithm of Stock and Watson (2002b) to estimate (17.6), and then estimate (17.7) and (17.8) by OLS on the estimated factors and idiosyncratic components.

The model of Giannone et al. (2008) and the model of Bańbura and Modugno (2014) have been used especially for *nowcasting*, where with the term "nowcasting" the literature refers to the prediction of the present such as, for example, the prediction of current quarter GDP growth rate. Actually, a large part of the literature on predicting economic activity in real-time with Factor models is dedicated to *nowcasting*. Indeed, due to publication delays of economic data, policy institutions, such as Central Banks and Ministries, are always forced to set their policies without knowing the current state of the economy, and, sometimes, even without knowing the recent past. Predicting the current/past state of the economy is therefore a crucial task. Given that this is such an important topic, we will discuss *nowcasting* empirical applications in the next Section.

17.3.1 Nowcasting

There exists a large literature on *nowcasting* with Dynamic Factor models. This literature was pioneered by Giannone et al. (2008), whose model was first implemented at the Board of Governors of the Federal Reserve Bank in the early 2000s. Since then, many authors have applied their approach on many different countries, and nowadays many Central Banks use this approach. A non-exhaustive list of papers that has used DFM for *nowcasting* is: Rünstler et al. (2009), Angelini et al. (2010), Angelini et al. (2011), Matheson (2010), Marcellino and Schumacher (2010), Barhoumi et al. (2010), Bańbura and Rünstler (2011), Aastveit and Trovik (2012), and D'Agostino et al. (2012) (for a review, see Bańbura et al. (2010) and Bańbura et al. (2011).

Figure 17.1 is taken from Giannone et al. (2008). Giannone et al. (2008) estimate the factors from a panel of about 200 monthly macroeconomic indicators, and their goal is to understand the effect on forecast accuracy of each release during the quarter. The left plot of Figure 17.1 shows the nowcast of US GDP conducted at the end of the first week of the second month of each quarter when the Survey of Professional forecasters is released. As we can see from this plot, the prediction obtained with the Factor model is qualitatively similar to the one obtained by the Survey of Professional Forecasters. The right plot, instead, reports the mean squared prediction error of the Factor model. From this plot we can see how the MSPE declines as more data becomes available. Moreover, from this plot it can be noticed that the release that has the largest (decreasing) impact on the MSPE is the "Mixed 2" block, which is the first block containing data relating to the current month (Philadelphia Business Outlook Survey).

Figure 17.2 shows the prediction of Euro Area GDP obtained with a Factor model similar to (17.3)-(17.5) estimated on a panel of 85 monthly indicators most of which are representative of the Euro Area economy Angelini et al. (2011). The upper/lower plot shows the prediction of GDP conducted seven/one months before GDP is released, that is, by using the terminology adopted by the literature, respectively a one-step ahead forecast and a *nowcast*. Comparing the two plots it is evident that the prediction obtained with the Factor model improves as more data are released. Conversely, the improvement obtained with the univariate bridge model is rather limited.

Finally, table 17.2, which reports results in Bańbura and Modugno (2014), reports Relative mean squared error of the prediction of Euro Area GDP obtained with model (17.6)-(17.8) estimated on a sample of 101 Euro Area variables. Results in Table 17.2 show how model (17.6)-(17.8) does better than univariate benchmarks, and how the prediction error decreases as more data become available.

To sum up, the *nowcasting* literature has shown that: (i) Dynamic Factor models produce *nowcasts* that outperform standard univariate benchmark such as random walk models, Autoregressive models, and Bridge models; (ii) Dynamic Factor models perform as well as institutional forecasts; and (iii) as more data pertaining the current quarter become available, the *nowcasting* error decreases monotonically, that is Dynamic Factor models are able to revise efficiently their prediction as new data are released.

Figure 17.1: Nowcasting US GDP
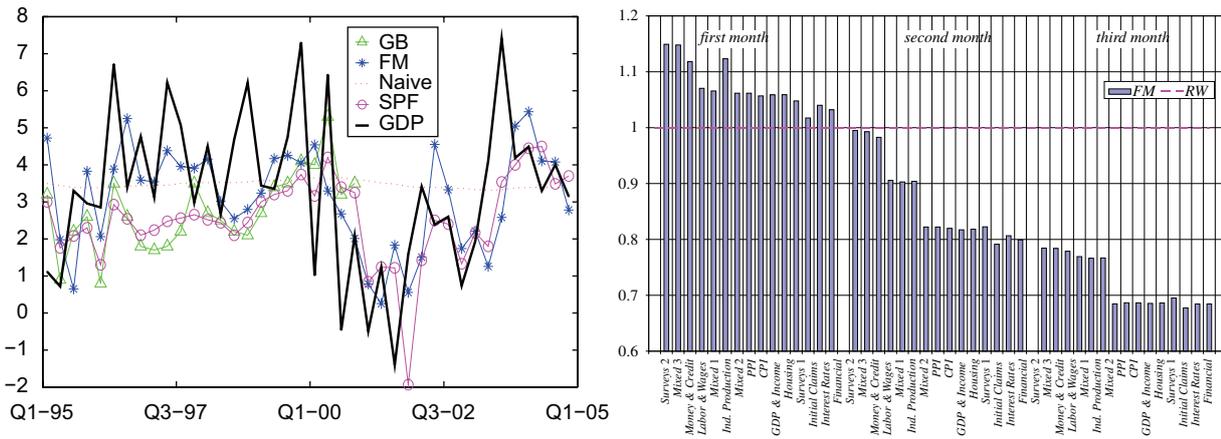
The left plot show the nowcast of GDP obtained with a Factor model (FM) conducted the end of the first week of the second month of each quarter. SPF = Survey of Professional Forecasts; GB = the Greenbooks nowcast, i.e. nowcast produced by the Board of Governors of the Federal Reserve

The right plot show the mean squared prediction error of the Factor model relative to the naive constant-growth model. The MSPE is computed in different point in time corresponding to the release of different blocks of data. First month stands for first month of the quarter, in the case of Q1 is January.

Source: This figure is taken from Giannone et al. (2008).

17.4 Constructing Business Cycle Indicators

In this Section, we will present two indicators, one for the US (the CFNAI) and one for the Euro Area (Eurocoin). We will then also briefly discuss construction of core inflation indexes.

The Chicago Fed National Activity Index

The Chicago Fed National Activity Index (CFNAI) is a monthly index constructed to provide an objective real-time measure of coincident economic activity for the US. The CFNAI is based on Stock and Watson (1999), which is one of the first papers proposing Large-Dimensional Factor models for forecasting. Specifically, let $\mathbf{x}_t = \mathbf{\Lambda} F_t + \boldsymbol{\xi}_t$, the CFNAI is simply F_t, and it is estimated as the first principal component extracted from a panel of 85 monthly indicators

The CFNAI is released at the end of each month, namely the value of the index for month T is released at the end of month $T + 1$ (e.g. on March 24 2014 it was released the value for February 2014). Since at the end of month $T + 1$ $\mathbf{x}_T$ is unknown for about one-third of the 85 series, when computing CFNAI in real-time there is a problem of missing observations, which is solved separately series by series by forecasting with an AR model up to the month in which the index is produced, see Brave and Butters (2014). Conversely, since all variables in $\mathbf{x}_t$ are monthly, in computing CFNAI there is no problem of mixed frequency

Figure 17.3 shows the time series of CFNAI-MA3 (the three months moving average of CFNAI) estimated at the end of April 2014. The estimated value for the CFNAI-MA3 in March is zero thus suggesting that growth was at its historical trend and that there is limited inflationary pressure over the coming year. Finally, the dash-dotted lines reported in Figure 17.3 are thresholds that often provides early indications of business cycle turning points and/or of changes in inflationary pressure Evans et al. (2002), Brave and Lichtenstein (2012). For example, in the left plot of Figure 17.3 values of CFNAI-MA3 below -0.7 following a period of expansion indicate an increased likelihood that a recession has begun, while when CFNAI-M3 is above -0.7 there is an

Figure 17.2: Nowcasting Euro Area GDP
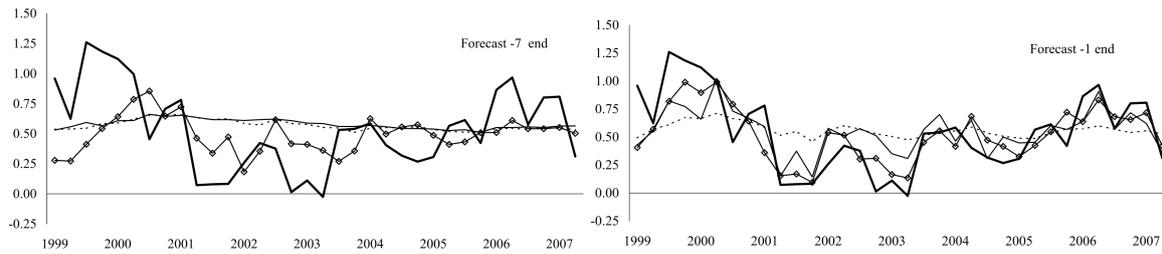
The upper/lower plot show the prediction of GDP conducted seven (left) and one (right) months before GDP is released. The thick line is actual GDP, the thin line with diamond markers is the prediction obtained with a Factor model, the thin line is the prediction obtained with a Bridge model based on Selected predictors, and finally the dotted line is the prediction obtained with a Bridge model based on All predictors.
Source: This figure is taken from Angelini et al. (2011).

increased likelihood that the recession has ended. Similarly, in the right plot of Figure 17.3 values of CFNAI-MA3 above 0.7 following more than two years into an economic expansion, indicates an increasing likelihood that a period of sustained increasing inflation has begun.

€-coin

Altissimo et al. (2001), Altissimo et al. (2010) developed a business cycle indicator for the Euro Area known as Eurocoin (€-coin), which is computed each month by the staff of the Banca d'Italia, and it is published also by the Centre for Economic Policy Research.

€-coin was constructed in order to have an assessment of economic activity that is free from short-run fluctuations, where: (i) with the term "economic activity" it is meant that the goal is to capture only those fluctuations due to the structural/macroeconomic shocks $\mathbf{u}_t$, i.e. the business cycle; and (ii) with the expression "free from short-run fluctuations" it is meant that the goal is not to estimate all the fluctuations induced by the common shocks, but only those producing medium-run and long-run fluctuations (i.e. larger than one year).

Let $\mathbf{x}_t$ be a large vector of month-on-month growth rates, and let y_t^Q be quarter-on-quarter GDP growth rate, it is assumed that $(\mathbf{x}_t, y_t^Q)$ have an approximate factor representation:

$$y_t^Q = \mathbf{\Lambda}^y \mathbf{F}_t^Q + \boldsymbol{\xi}_t^y \tag{17.1}$$

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{17.2}$$

where $\mathbf{F}_t^Q = (1 + L + L^2)^2 \mathbf{F}_t$ are the monthly factors transformed in quarterly figures (see Annex 17.A.4). Further, let c_t^Q be medium-run to long-run GDP growth,[8] and let $\mathbf{f}_t^Q = (1 + L + L^2)^2 \mathbf{f}_t$ be the quarterly smoothed factors, i.e. the factors that captures medium to long-run fluctuations,[9] then €-coin is the linear projection of c_t^Q onto the space spanned by $\mathbf{f}_t^Q$, namely

$$\text{€-coin} = Proj(c_t^Q / \mathbf{f}_t^Q) = \bar{y}_t^Q + \mathbf{\Sigma}^{cfQ}(\mathbf{\Sigma}^{fQ})^{-1}\mathbf{f}_t^Q$$

---

[8] Medium-run to long-run GDP growth is estimated by applying to y_t^q a band-pass filter Baxter and King (1999) that retains only frequencies between $-\pi/6$ and $\pi/6$, that is fluctuations larger than one year.

[9] The smoothed factors $\mathbf{f}_t$ are estimated as the generalized principal components (see Annex 17.A.2) of $\mathbf{x}_t$ relative to the couple $(\widehat{\mathbf{\Sigma}}^\phi, \widehat{\mathbf{\Sigma}}^\chi + \widehat{\mathbf{\Sigma}}^\xi)$, where $\phi_t = w(L)\chi_t$ is the part of the common component related to fluctuations larger than one year, and $\widehat{\mathbf{\Sigma}}^\phi = \int_{-\pi/6}^{\pi/6} \widehat{\mathbf{S}}^\chi(\theta)$.

Table 17.2: Root Mean Squared Prediction Error of model (17.6)- (17.8)

	Factor Models		Benchmarks	
	RW	AR(1)	AR	RW
M6	0.28	0.29	0.33	0.32
M5	0.27	0.26	0.32	0.32
M4	0.26	0.26	0.32	0.32
M3	0.24	0.24	0.32	0.32
M2	0.24	0.23	0.27	0.31
M1	0.25	0.24	0.27	0.31
M0	0.21	0.20	0.27	0.31

This table reports mean squared prediction error obtained at different points in time, namely at the end of the first/second/third month of the previous quarter (M6/M5/M4), at the end of the first/second/third month of the current quarter (M3/M2/M1), and at the end of the first month of the next quarter. To give an example, for Q2 M6 is January, M3 is March, and M0 is July. The evaluation sample is 2000:Q1 to 2007:Q4.

This table reports results for two different specification of model (17.6)-(17.8), in the first the idiosyncratic components is assumed to be white noise, while in the second the idiosyncratic component is assumed to follow an AR(1) model. Similarly, this table reports results for two benchmark models, AR model and a RW model.

Source: Bańbura and Modugno (2014).

where $\bar{y}_t^Q$ is the average quarterly GDP growth (i.e. the mean of y_t^Q), $\mathbf{\Sigma}^{cf}$ is the row vector containing the covariances between c_t^Q and $\mathbf{f}_t^Q$, and $\mathbf{\Sigma}^f$ is the covariance matrix of $\mathbf{f}_t^Q$.[10]

The value of €-coin for month T is calculated and released at the end of month T. Similarly to CFNAI, since all the variables in $\mathbf{x}_t$ are monthly indicators, there is no mixed frequency problem when computing €-coin in real/time. However, when computing the index not all monthly indicators are available for month T. Differently from CFNAI, though, the problem of missing values is solved by realigning the series: let x_{it}^* be the original series, and suppose that the last available value at the time the index is calculated is $x_{iT-k_i}^*$, then to realign means that $x_{iT} = x_{iT-k_i}^*$.

Figure 17.4 shows the time series of €-coin computed at the end of the April 2014, together with GDP growth. As we can see, with the exception of 2008Q4 and 2009Q1 when €-coin overestimated GDP growth, overall the indicator tracks well Euro Area GDP growth. The estimated value for €-coin in April 2014 is 0.39 (+0.01 compared to March 2014) thus suggesting no changes in the Euro Area business cycle as of end of April 2014.

Core Inflation Indexes

Finally, Factors models have been used not only for constructing business cycles indicators, but also for constructing core inflation indexes, where the goal when constructing such index is to remove from a given measure of inflation the effect of transitory shocks, of measurement errors, and of sectorial shocks. Cristadoro et al. (2005) construct a core inflation index for the Euro Area by using techniques similar to that used by

---

[10]Once $\mathbf{f}_t$ is estimated, estimation of $\mathbf{\Sigma}^{fQ}$ is trivial. Conversely, $\mathbf{\Sigma}^{cfQ}$ is estimated as follows: (i) estimate the cross-covariances between y_t^Q and $\mathbf{f}_t^Q$; (ii) from these cross-covariances, estimate the cross-spectrum between y_t^Q and $\mathbf{f}_t^Q$, $\widehat{\mathbf{S}}^{yf}(\theta)$; (iii) estimate $\mathbf{\Sigma}^{cfQ}$ as $\widehat{\mathbf{\Sigma}}^{cfQ} = \int_{-\pi/6}^{\pi/6} \widehat{\mathbf{S}}^{yfQ}(\theta)$.

**Figure 17.3:** CFNAI Real-Time Estimation
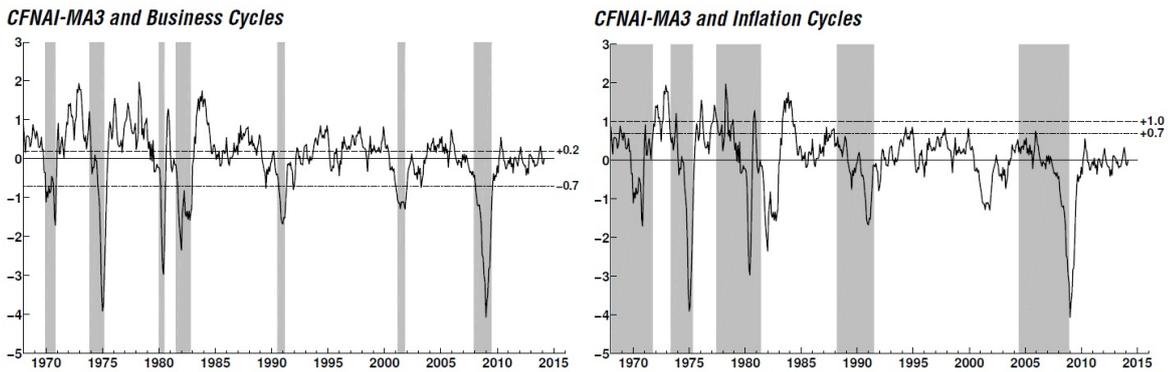
The black line in the left and in the right plot is the CFNAI-MA3 published in the CFNAI News Release of April 21, 2014. The shaded area in the left plot are periods of recession as identified by the NBER, while the shaded area in the right plot are periods of substantial inflation increases as identified inn the pdf "Background on the Chicago Fed National Activity Index".[a] Finally, the dash-dotted lines in the left and in the right plot are thresholds values that often provides early indications of business cycle turning points or of changes in in inflationary pressure Evans et al. (2002), Brave and Lichtenstein (2012).
Source: CFNAI News Release April 21, 2014.[b]

---

[a]Available at https://chicagofed.org/~/media/publications/cfnai/background/cfnai-background-pdf.pdf
[b]See also http://www.chicagofed.org/webpages/publications/cfnai/.

Figure 17.4: €-coin Real-Time Estimation
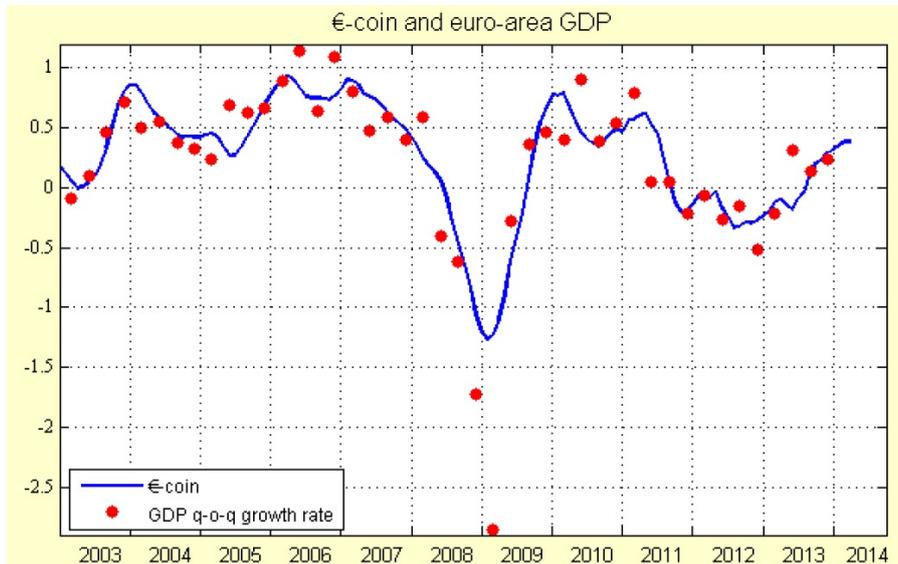
The blue line is €-coin published on the €-coin News Release 30 April 2014.
The red dots are Euro Area quarter-on-quarter GDP growth.
Source: €-coin News Release 30 April 2014,
see also http://eurocoin.bancaditalia.it.

Altissimo et al. (2010), while Giannone and Matheson (2007) use the same model of Cristadoro et al. (2005) to construct a core index for New Zealand.

In a nutshell, the core inflation index of Cristadoro et al. (2005) is constructed as that part of the common

component of the HICP which captures fluctuations larger than one year. Let $\mathbf{x}_t$ be a large panel of monthly variables that have an approximate factor representation so that $\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$. The common component $\boldsymbol{\chi}_t$ can be decomposed in a part which includes only fluctuations larger than one year, $\boldsymbol{\phi}_t$, and a part that includes only fluctuations shorter than one year, $\boldsymbol{\psi}_t$, $\boldsymbol{\chi}_t = \boldsymbol{\phi}_t + \boldsymbol{\psi}_t$. Suppose that HICP is the first variable in the panel, x_{1t}, then the core index of Cristadoro et al. (2005) is $x_{1t} = \phi_{1t}$.

17.5 Large-Dimensional Databases

In the Introduction we have argued that models that are able to include a large number of predictors are believed to improve the forecast. Indeed, there is lots of evidence, some of which we reported in the previous Sections, that Large-Dimensional Factor models are extremely powerful in predicting economic activity. Moreover, in Section 17.2.2 we have shown that a large number of variables (ideally an infinite number) is necessary to consistently estimate Approximate Dynamic Factor models. In this Section we will discuss two issues related with large-dimensional databases that received attention in the literature. We will start in Section 17.5.1 by questioning the usefulness of large-dimensional databases for forecasting, and then in Section 17.5.2 we will discuss some results in the literature, that show how when constructing a database one should be careful in not including *noisy* variables.

17.5.1 Large- *vs.* Small-Dimensional Databases

In this section we discuss a branch of the literature which has investigated whether there are gains in terms of forecasting by first selecting a limited number of variables, and then estimating the Factor models on the small database containing the selected variables. The literature has suggested three possible strategies for variable selection:

- variable selection by statistical methods is used by Bai and Ng (2008a) and Camacho and Perez-Quiros (2010) among others. Specifically, Bai and Ng (2008a) suggests selecting only those variables that are really informative for forecasting the target variable with the LARS algorithm, while Camacho and Perez-Quiros (2010) recommends first selecting a core group of variables, and then evaluating if other possible predictors are useful. This strategy is also adopted by Schumacher (2010) and Marcellino et al. (2013) for example.

- variable selection with "economic judgment" essentially amount to include only aggregate variables. This strategy is adopted by Bańbura et al. (2010), Bańbura et al. (2011) and Luciani (2014) for example.

- variable selection by market operators – pioneered by Bańbura et al. (2012) and followed also by Luciani and Ricci (2014) – consists of considering only those variables that are followed closely by the market. The rationale of this approach is that market participants can be viewed as nowcasters. Market participants monitor macroeconomic data, form their expectations on current and future fundamentals of a given country, and, also, based on these data/expectations allocates their investments. Since it is realistic to assume that they know what are the relevant series to monitor in order to form appropriate expectations on current GDP, it make sense to include in the model only those variables that are followed by them.

In a nutshell, the conclusion of this literature is that, unless one needs to comment on many data releases, as it can be the case for example of a policy report, there is no need of a large database when forecasting with Dynamic Factor Model as long as the variables on which the model is estimated are appropriately selected with one of the methods just described. For example, Table 17.1 reports results from Bańbura et al. (2011) which show how in terms of forecasting accuracy estimating a Dynamic Factor model as (17.6)-(17.8) on a

database with or without disaggregated variables makes no difference.[11] As another example, Table 17.2 reports results from Bańbura et al. (2012) which show how the MSPE of a Dynamic Factor model estimated on 20 monthly indicators all targeted by market operators, relative to the MSPE of a Random Walk model (0.64) is comparable to the performance of a large-dimensional Dynamic Factor model such as that reported in the right plot of Figure 17.1.

Table 17.1: MSPE: Small Database: Economic judgment

	SDFM	LDFM	RW	AR
M6	0.27	0.27	0.32	0.33
M5	0.26	0.25	0.32	0.32
M4	0.25	0.25	0.32	0.32
M3	0.23	0.23	0.32	0.32
M2	0.21	0.21	0.31	0.27
M1	0.20	0.21	0.31	0.27
M0	0.18	0.20	0.31	0.27

This table reports mean squared prediction error obtained at different points in time, namely at the end of the first/second/third month of the previous quarter (M6/M5/M4), at the end of the first/second/third month of the current quarter (M3/M2/M1), and at the end of the first month of the next quarter. To give an example, for Q2 M6 is January, M3 is March, and M0 is July.

Column SDFM reports the MSPE of a Dynamic Factor model estimated on a small database including only aggregate variables, column LDFM reports the MSPE of a Dynamic Factor model estimated on a large database including also disaggregated variables, while columns RW and AR report, respectively, the MSPE for a random walk and an autoregressive model.

Source: Bańbura et al. (2011)

17.5.2 How to construct the database

In this Chapter we have discussed Large Dimensional Factor models, meaning factor models estimated on large dataset. The question is: in practice large dataset what does it mean? This is a question of general interest for factor analysis and of particular importance for practitioners. The construction of a database is, indeed, a practical problem for which there is no recipe, and one which practitioners unavoidably face when working with factor models. In principle, given that factor estimates are consistent as $n \to \infty$, including all available variables is a natural choice. Is this the right choice? How many, and which variables do we have to include in the analysis? And, are there variables that are worth excluding from the analysis?

A number of papers discuss whether when forecasting with factor models it is always useful to increase the size of the database. Consider, for example, Producer Price Indexes (PPI). From the FRED database of the Federal Reserve Bank of St. Louis it is possible to easily download more than fifteen PPIs. Is it a good strategy to include them all in the database, or would it be better to select just one or a few of them? The literature suggests that the correct answer is: "we should include one or a few of them".

In a nutshell, the literature has shown that if adding an extra variable does not add information about the factors, but instead it simply adds an extra cross-correlation among idiosyncratic errors, i.e. noise, then (i) the estimates of the factors will deteriorate, and similarly (ii) tests and criteria for the number of factors will fail, thus (iii) the forecasting performance of the model deteriorates. More precisely, the simulations of Boivin and

---

[11]This result is confirmed by Barhoumi et al. (2010) who show that medium size models (i.e. including 10-30 variables) perform equally well in forecasting than large size models (about 100 variables). Furthermore, Luciani (2014) shows that while aggregate variables are enough to produce a good forecast of GDP, when forecasting more disaggregated variables, sectoral information matters.

Table 17.2: MSPE: Small Database: Market operators

Q	M	D	DFM	Bridge	RW	SPF
0	1	7	0.57	0.66	0.66	
0	1	14	0.58	0.65	0.66	
0	1	21	0.54	0.65	0.66	
0	1	28	0.54	0.64	0.66	
0	2	7	0.52	0.62	0.66	
0	2	14	0.51	0.62	0.66	0.49
0	2	21	0.51	0.61	0.66	
0	2	28	0.50	0.60	0.66	
0	3	7	0.49	0.59	0.66	
0	3	14	0.47	0.59	0.66	
0	3	21	0.46	0.59	0.66	
0	3	28	0.45	0.58	0.66	
1	1	7	0.45	0.58	0.66	
1	1	14	0.45	0.58	0.66	
1	1	21	0.44	0.57	0.66	
1	1	28	0.42	0.57	0.66	

This table shows the root mean squared prediction error of a Dynamic Factor model estimated on 20 monthly indicators all targeted by market operators ("DFM"), of the average produced by the bridge equations ("Bridge"), of a random walk model ("RW"), and of the Survey of Professional Forecasters ("SPF"). The mean squared prediction errors are computed four times per month for four consecutive months (from the first month of the current quarter, $Q = 0$ and $M = 1$ to the first month of the following quarter, $Q = 1$ and $M = 1$).
Source: Bańbura et al. (2012)

Ng (2006) show that, as the cross-correlation among the idiosyncratic errors increases, the estimation and forecasting performance of the model deteriorates. Luciani (2014) show that tests and criteria for determining the numbers of factors are extremely unreliable when the database is poorly constructed, often resulting in either an underestimation or an overestimation of the number of factors. Onatski (2012) shows that if the explanatory power of the factors does not strongly dominate the explanatory power of the idiosyncratic terms, meaning that pervasive and non-pervasive shocks cannot be distinguished clearly, then the principal component estimator is inconsistent. In addition, the simulation results of Bai and Ng (2008b) confirm that the factor estimates can be severely compromised in such situations.

In summary, in constructing the database, the researcher should try to include enough variables to represent properly the economy he/she is analyzing, but not too many variables. In particular, including variables that are too similar from one another is a bad strategy since it just deteriorates the performances of the model.

17.6 Conclusions

In this Chapter we review the literature on Large-Dimensional Dynamic Factor Models for real-time applications. We first present the Dynamic Factor model, the implications of using large-dimensional databases, and the challenges of real-time applications due to non-synchronous releases of different variables, and mixed frequencies. We then discuss how the literature has solved these problems, and we introduce the most successful specification of the model that have been used in real-time for constructing business cycle indicators, as well as for predicting economic activity. Then, we present numerous empirical applications (in particular on *nowcasting*) in order to show the usefulness of the Dynamic Factor model for real-time applications, and in

particular how the model is able to correctly revise its predictions as new information become available. Lastly, we discuss recent literature that has questioned the usefulness of large-dimensional databases.

Annex

17.A Annex

17.A.1 Pervasive and non-pervasive shocks

As we explained in Section 17.2.1 the difference between common and idiosyncratic shocks is that the former are pervasive while the latter are non-pervasive. Formally, let us denote the covariance matrix of $\mathbf{x}_t$ as $\mathbf{\Sigma}^x = \mathbb{E}(\mathbf{x}_t \mathbf{x}_t')$, and its j-largest eigenvalue as μ_j^x. Similarly, let us denote the spectral density matrix of $\mathbf{x}_t$ at a generic frequency θ as $\mathbf{S}(\theta) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\theta} \mathbf{\Sigma}^x(h)$, where $\Sigma^x(h) = \mathbb{E}(\mathbf{x}_t \mathbf{x}_{t-h}')$, and its j-largest eigenvalue as $\lambda_j^x(\theta)$. Finally, let us denote as c a constant different from zero, then:

Assumption 1. *Pervasiveness of common shocks: The common shocks have a non-negligible effect on all the variables meaning that the eigenvalues of the spectral density matrix of the common component grow with n almost everywhere between $-\pi$ and π. Formally: $\lim_{n \to \infty} \frac{1}{n} \lambda_j^\chi(\theta) = c$, for θ a.e. $[-\pi, \pi]$.*

Assumption 2. *Non-pervasiveness of idiosyncratic shocks: The idiosyncratic shocks affect only a limited number of variables, meaning that the filter $\mathbf{D}(L)$ defined in (17.7) is such that the eigenvalues of the covariance matrix of the idiosyncratic component are bounded for any n: $\mu_1^\xi = c \leq \infty$.*

Since the common and the idiosyncratic shocks are uncorrelated at all leads and lags, the spectral density matrix of the data can be written as: $\mathbf{S}^x(\theta) = \mathbf{S}^\chi(\theta) + \mathbf{S}^\xi(\theta)$. Now, from Weyl Inequality we have that $\lambda_j^\chi(\theta) + \lambda_n^\xi(\theta) \leq \lambda_j^x(\theta) \leq \lambda_j^\chi(\theta) + \lambda_1^\xi$, which implies that $\lim_{n \to \infty} \lambda_j^x(\theta) = \infty$ if $j = 1 \ldots q$ and $\lim_{n \to \infty} \lambda_j^x(\theta) = c \leq \infty$ for $j \geq q + 1$. That is, asymptotically, the q largest eigenvalues of the spectral density matrix of the data diverge to infinity, while the other are bounded, or, in other words, as the number of variables get larger and larger, since the information about χ_t accumulates as we add variables, the common components dominates. Indeed, Forni and Lippi (2001) prove also that if the q-th largest eigenvalues of $\mathbf{S}^x(\theta)$ diverges, while the $(q + 1)$-th is bounded then $\mathbf{x}_t$ admits a representation such as (17.1)-(17.3), and moreover the decomposition into common and idiosyncratic components is unique, meaning that the number of factors q and the common and idiosyncratic components are uniquely identified, thus a representation with a different number of shocks is not possible.

When working with the restricted model (17.4)-(17.7) it is also necessary to assume that the eigenvalues of the covariance matrix of the common component grow with n ($\lim_{n \to \infty} \frac{1}{n} \mu_j^\chi = c$). An alternative way of making explicit this assumption is by assuming that $\lim_{n \to \infty} \frac{1}{n} \mathbf{\Lambda}' \mathbf{\Lambda} = \bar{\mathbf{\Lambda}}$, where $\bar{\mathbf{\Lambda}}$ is a full rank matrix, which ensures that the factor loadings are sufficiently heterogeneous so that the factors affect most or all of the variables (see Stock and Watson , 2011).

Independence of common and the idiosyncratic shocks, allows to decompose the covariance matrix of the data can be written as $\mathbf{\Sigma}^x = \mathbf{\Sigma}^\chi + \mathbf{\Sigma}^\xi$. Again, by using the Weyl Inequality we can show that asymptotically, the r largest eigenvalues of $\mathbf{\Sigma}^x$ diverge to infinity, while the other are bounded, which is why in Section 17.2.2 we suggest that the weights $\mathbf{w}$ to be used for estimating the factors should be such that they maximize the variance of the cross-sectional averages.

Finally, notice that an alternative way of looking at the problem of factor estimation is to require that the estimated factors ($\widehat{\mathbf{F}}_t$) and the estimated loadings ($\widehat{\mathbf{\Lambda}}$) should be such that they maximize the variance of the

common component. This leads to solve the least square problem:

$$\min_{\{\mathbf{F}_t\}_{t=1}^T, \boldsymbol{\Lambda}} \frac{1}{nT} \sum_{t=1}^T (\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_t)'(\mathbf{x}_t - \boldsymbol{\Lambda}\mathbf{F}_t) \qquad \text{s.t.} \quad \frac{1}{n}\boldsymbol{\Lambda}'\boldsymbol{\Lambda} = \mathbf{I}. \qquad (17.A.1)$$

The solution to the problem in (17.A.1) is:

$$\widehat{\boldsymbol{\Lambda}} = \mathcal{V}_r \mathcal{D}_r^{1/2}$$

$$\widehat{\mathbf{F}}_t = \frac{1}{n}\widehat{\boldsymbol{\Lambda}}'\mathbf{x}_t,$$

which is equivalent to finding the first r principal components of $\mathbf{x}_t$.

17.A.2 Generalized Principal Components

Let $\mathbf{x}_t$ be as in (17.4)–(17.7). Denote $\widehat{\mathbf{S}}^\chi(\theta)$ and $\widehat{\mathbf{S}}^\xi(\theta)$ as the estimated spectral density matrices of, respectively, the common and the idiosyncratic component obtained with the method of dynamic principal components. Then the covariance matrices of χ_t and ξ_t can be consistently estimated as the inverse Fourier transform of $\widehat{\mathbf{S}}^\chi(\theta)$ and $\widehat{\mathbf{S}}^\xi(\theta)$, namely: $\widehat{\boldsymbol{\Sigma}}^\chi = \int_{-\pi}^\pi \widehat{\mathbf{S}}^\chi(\theta)$, and $\widehat{\boldsymbol{\Sigma}}^\xi = \int_{-\pi}^\pi \widehat{\mathbf{S}}^\xi(\theta)$.

Denote $\widehat{\mathbf{Z}}$ as the $n \times r$ matrix containing the first normalized r eigenvectors of $\widehat{\boldsymbol{\Sigma}}^\chi(\widehat{\boldsymbol{\Sigma}}^\xi)^{-1}$, that is $\widehat{Z}_j$ solves $\widehat{\boldsymbol{\Sigma}}^\chi Z_j = v_j \widehat{\boldsymbol{\Sigma}}^\xi Z_j$ subject to $Z_j' \widehat{\boldsymbol{\Sigma}}^\xi Z_j = 1$, then the common factors and the factor loadings are consistently estimated as:

$$\widehat{\mathbf{F}}_t = \widehat{\mathbf{Z}}'\mathbf{x}_t \qquad (17.A.2)$$

$$\boldsymbol{\Lambda} = \widehat{\boldsymbol{\Sigma}}^\chi \widehat{\mathbf{Z}} \left(\widehat{\mathbf{Z}}'\widehat{\boldsymbol{\Sigma}}^\chi\widehat{\mathbf{Z}}\right)^{-1} \qquad (17.A.3)$$

that is the common factors are the generalized principal components of $\mathbf{x}_t$ relative to the couple $(\widehat{\boldsymbol{\Sigma}}^\chi, \widehat{\boldsymbol{\Sigma}}^\xi)$, while the loadings are the linear projection of the static factors onto the space spanned by $\mathbf{x}_t$. Alternatively, (17.A.2) can be seen as the principal component estimator of $\mathbf{F}_t$ on $\check{\mathbf{x}}_t = (\widehat{\boldsymbol{\Sigma}}^\xi)^{-1/2}\mathbf{x}_t$.

17.A.3 The Expectation-Maximization Algorithm

In this Section we will discuss two EM algorithms proposed in the literature to estimate Dynamic Factor models when there are mixed frequencies and missing values. The main idea of the EM algorithm is to by-pass the problem of missing observations by maximizing the likelihood as if there were no missing observations and no mixed frequencies. Concretely, this amounts to iterate between an expectation step (E-step) in which missing values are filled and mixed frequencies are accounted for, and a maximization step (M-step) in which the expected likelihood is maximized.

Stock and Watson (2002b)

In Section 17.2.2 we discussed estimation of $\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \xi_t$, and we derived the estimator for $\mathbf{F}_t$ and $\boldsymbol{\Lambda}$ as a solution of (17.A.1), which is equivalent to Principal Components. Starting from the observation that the function in (17.A.1) is proportional to the log-likelihood of $\mathbf{x}_t = \boldsymbol{\Lambda}\mathbf{F}_t + \xi_t$ under the assumption that $\mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\Lambda}\mathbf{F}_t, \mathbf{I}_n)$, and hence in this case the Principal Component estimator derived in section 17.2.2 is the Gaussian maximum likelihood estimator. Stock and Watson (2002b) derive an EM algorithm for estimating $\mathbf{F}_t$ and $\boldsymbol{\Lambda}$.

Let:

(i) $\mathbf{x}_t$ be the original panel of variables containing monthly and quarterly variables, and missing values as well,

(ii) $\tilde{\mathbf{x}}_t$ be the panel of variables without missing values and with all the variables at the higher frequency, and

(iii) let $\widehat{\mathbf{F}}_t^{(j)}$ and $\widehat{\mathbf{\Lambda}}^{(j)}$ be the estimates obtained at the j-th iteration.

Then, the algorithm at the $j + 1$-th iteration works as follows:

– E-Step:construct the panel $\tilde{\mathbf{x}}_t$, that is fill the missing values and account for the mixed frequency as follows:

– missing values:$\tilde{x}_{it} = x_{it}$ if observed, $\tilde{x}_{it} = \widehat{\mathbf{\Lambda}}_i^{(j)}\widehat{\mathbf{F}}_t^{(j)}$ otherwise.

– mixed frequencies: $\tilde{x}_{it} = \widehat{\mathbf{\Lambda}}_i^{(j)}\widehat{\mathbf{F}}_t^{(j)} + \widehat{\xi}_{it}^{(j)}$, where by exploiting the approximation of Mariano and Murasawa (2003) explained in 17.A.4 we have $\widehat{\xi}_{it}^{(j)} = x_{it}^q - \widehat{\mathbf{\Lambda}}_i^{(j)}(\widehat{\mathbf{F}}_\tau^{(j)} + 2\widehat{\mathbf{F}}_{\tau-1}^{(j)} + 3\widehat{\mathbf{F}}_{\tau-2}^{(j)} + 2\widehat{\mathbf{F}}_{\tau-3}^{(j)} + \widehat{\mathbf{F}}_{\tau-4}^{(j)})$ where $\tau = 3$ for $t = 1, 2, 3$, $\tau = 6$ for $t = 4, 5, 6$, and so on.

– M-Step: run principal components on $\tilde{\mathbf{x}}_t$ to get new estimates of the factors and of the loadings, $\widehat{\mathbf{F}}_t^{(j+1)}$ and $\widehat{\mathbf{\Lambda}}^{(j+1)}$.

– Stopping rule: The algorithm is stopped when the difference in the sum of the squared idiosyncratic components –i.e. the function in (17.A.1)– between iterations reaches a desired level of convergence.

– Initialization: The algorithm can be initialize either by filling missing values, and by assigning monthly values to the quarterly data, or, in case there exists a subset of the original dataset that constitutes a balanced panel, initial values can be obtained from the balanced dataset.

Bańbura and Modugno (2014)

Bańbura and Modugno (2014) modify the EM algorithm derived by Watson and Engle (1983) in order to estimate the parameters of the Dynamic Factor model on a dataset with an arbitrary pattern of missing data. The model of Bańbura and Modugno (2014) that we discuss in Section 17.3 is:

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \boldsymbol{\xi}_t \tag{17.6}$$

$$\mathbf{F}_t = \mathbf{A}\mathbf{F}_{t-1} + \mathbf{u}_t \tag{17.7}$$

$$\xi_{it} = \alpha_i\xi_{it-1} + e_{it} \qquad i = 1, \dots, n. \tag{17.8}$$

Let $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, and assume that $\alpha_i = 0$ and $\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi})$ where $\mathbf{\Psi}$ is a diagonal matrix with diagonal elements $\psi_1, \psi_2, \dots, \psi_n$ –that is the idiosyncratic components are assumed to be cross-correlated in the sense of Assumption 1, but for estimation is imposed $\mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \mathbf{\Psi}$.Collect all parameters $\{\mathbf{\Lambda}, \mathbf{A}, \mathbf{Q}, \mathbf{\Psi}\}$ in the vector $\boldsymbol{\theta}$, and suppose that the algorithm was run $j - 1$ times, then the j-th iteration works as follows:

– E-step: given $\boldsymbol{\theta}^{(j-1)}$, estimate $\widehat{\mathbf{F}}_t^{(j)}$ with the Kalman Filter/Smoother modified for accounting for both missing values and mixed frequency as follows:

– missing values: missing values are handled by modifying the variance of the idiosyncratic component, namely as $\mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \tilde{\mathbf{\Psi}}$, where $\tilde{\psi}_{it} = \psi_{it}$ if x_{it} is observed, and $\tilde{\psi}_{it} = \infty$ otherwise.

– mixed frequencies: quarterly and monthly variables are handled by using the state-space representation (17.A.15), see 17.A.4.

– <u>M-step</u>: estimate the parameters in θ as if $\mathbf{F}_t$ was observed, that is

$$\widehat{\mathbf{\Lambda}}^{(j)} = \left[\sum_{t=1}^{T} \mathbb{E}(\mathbf{x}_t \mathbf{F}'_t) \right] \left[\sum_{t=1}^{T} \mathbb{E}(\mathbf{F}_t \mathbf{F}'_t) \right]^{-1} \tag{17.A.4}$$

$$\widehat{\mathbf{A}}^{(j)} = \left[\sum_{t=2}^{T} \mathbb{E}(\mathbf{F}_t \mathbf{F}'_{t-1}) \right] \left[\sum_{t=2}^{T} \mathbb{E}(\mathbf{F}_{t-1} \mathbf{F}'_{t-1}) \right]^{-1} \tag{17.A.5}$$

Note that (17.A.4)-(17.A.5) resemble OLS formulas for $\widehat{\mathbf{\Lambda}}^{(j)}$ and $\widehat{\mathbf{A}}^{(j)}$, the only difference being that since $\mathbf{F}_t$ is not observed, it is necessary to take expectations of it. In taking expectations, we have to consider that the true factors $\mathbf{F}_t$ are equal to the estimated factors plus an estimation error: $\mathbf{F}_t = \widehat{\mathbf{F}}_t + \boldsymbol{\eta}_t$, where $\mathbb{E}(\boldsymbol{\eta}_t) = 0$, $\mathbb{E}(\mathbf{x}_t \boldsymbol{\eta}'_t) = 0$, and $\mathbb{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}'_t) = \mathbf{\Sigma}_\eta^{t/t}$, which is an output of the Kalman filter/smoother. This implies that in practice $\widehat{\mathbf{\Lambda}}^{(j)}$ and $\widehat{\mathbf{A}}^{(j)}$ are estimated as follows:

$$\widehat{\mathbf{\Lambda}}^{(j)} = \left[\sum_{t=1}^{T} \mathbf{x}_t \widehat{\mathbf{F}}_t^{(j)\prime} \right] \left[\sum_{t=1}^{T} \widehat{\mathbf{F}}_t^{(j)} \widehat{\mathbf{F}}_t^{(j)\prime} + \mathbf{\Sigma}_\eta^{t/t} \right]^{-1} \tag{17.A.6}$$

$$\widehat{\mathbf{A}}^{(j)} = \left[\sum_{t=2}^{T} \widehat{\mathbf{F}}_t^{(j)} \widehat{\mathbf{F}}_{t-1}^{(j)\prime} \right] \left[\sum_{t=2}^{T} \widehat{\mathbf{F}}_{t-1}^{(j)} \widehat{\mathbf{F}}_{t-1}^{(j)\prime} + \mathbf{\Sigma}_\eta^{t-1/t-1} \right]^{-1} \tag{17.A.7}$$

from which we can estimate $\mathbf{Q}$ and $\mathbf{\Psi}$ as follows:

$$\widehat{\mathbf{\Psi}}^{(j)} = diag \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{x}_t - \widehat{\mathbf{\Lambda}}^{(j)} \widehat{\mathbf{F}}_t^{(j)} \right) \left(\mathbf{x}_t - \widehat{\mathbf{\Lambda}}^{(j)} \widehat{\mathbf{F}}_t^{(j)} \right)' \right\} \tag{17.A.8}$$

$$\widehat{\mathbf{Q}}^{(j)} = \frac{1}{T} \sum_{t=2}^{T} \left(\widehat{\mathbf{F}}_t^{(j)} - \widehat{\mathbf{A}}^{(j)} \widehat{\mathbf{F}}_{t-1}^{(j)} \right) \left(\widehat{\mathbf{F}}_t^{(j)} - \widehat{\mathbf{A}}^{(j)} \widehat{\mathbf{F}}_{t-1}^{(j)} \right)' \tag{17.A.9}$$

Finally, in order to take into account missing values only the available data are used in the calculations (17.A.6) and (17.A.8), meaning that, for the generic variable i, (17.A.6) and (17.A.8) are computed only for those t for which x_{it} is observed.

– <u>Stopping rule</u>: The algorithm is run until the increase in the likelihood (which can be obtained from the Kalman filter) between two consecutive steps is small.

– <u>Initialization:</u>The algorithm is initialized by estimating $\mathbf{F}_t$ and $\mathbf{\Lambda}$ by principal components, and $\mathbf{A}$ by OLS. This initialization is performed either by filling missing values, or from a balanced subset of the original dataset.

Bańbura and Modugno (2014) discuss also estimation when $\alpha_i \neq 0$ and $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi})$. In this setting, once the state-space representation is properly modified, model (17.6)-(17.8) can be estimated using the EM algorithm just described. Bańbura and Modugno (2014) suggest to modify the state-space representation by increasing the number of state variables, i.e. by adding the idiosyncratic component $\boldsymbol{\xi}_t$ to the state vector, and by augmenting the observation equation with an ad-hoc error with tiny variance, $\zeta_{it} \sim \mathcal{N}(0, \kappa)$, where κ is small. With these two modifications the state-space representation becomes:

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{F}_t \\ \boldsymbol{\xi}_t \end{pmatrix} + \boldsymbol{\zeta}_t \tag{17.A.10}$$

$$\begin{pmatrix} \mathbf{F}_t \\ \boldsymbol{\xi}_t \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & diag(\alpha_1, \dots, \alpha_n) \end{pmatrix} \begin{pmatrix} \mathbf{F}_{t-1} \\ \boldsymbol{\xi}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t \\ \mathbf{e}_t \end{pmatrix} \qquad \begin{pmatrix} \mathbf{u}_t \\ \mathbf{e}_t \end{pmatrix}. \tag{17.A.11}$$

17.A.4 State-Space representation with mixed frequencies

In this section we will explain how mixed frequencies are handled within the Kalman Filter. The material covered here is taken from Bańbura and Modugno (2014) and the interested reader is referred to these references for more details. Actually, here we will discuss the Dynamic Factor model with monthly and quarterly variables, that is the higher frequency we will consider is monthly. The interested reader is referred to Modugno (2014) for mixed frequency Dynamic Factor models with highest sampling frequency higher than monthly.

In mixed frequency Dynamic Factor models, the model is specified and estimated at the monthly frequency, but quarterly variables are included by constructing partially observed monthly counterparts. Practically, quarterly series are treated as monthly series with missing observations, and by assigning the quarterly observation to the 3rd month. The problem then is how to relate the monthly factors and the monthly idiosyncratic components with quarterly series. To answer this question it is first necessary to understand what is the relationship between monthly and quarterly growth rates.

Let t denotes months, and let Y_t^Q be the log-level of a quarterly variable. The quarter-on-quarter growth rate is then equal to $Y_t^Q - Y_{t-3}^Q$ and we will denote it as y_t^Q. Then, let Y_t^M be the monthly log-level of Y, and let $y_t^M = Y_t^M - Y_{t-1}^M$ be the month-on-month growth rate. We want to understand what would is the relationship between y_t^M and y_t^Q.

In order to link y_t^Q and y_t^M we follow Mariano and Murasawa (2003), and we make use of the approximation

$$Y_t^Q \approx Y_t^M + Y_{t-1}^M + Y_{t-2}^M \tag{17.A.12}$$

By substituting (17.A.12) into the equation $y_t^Q = Y_t^Q - Y_{t-3}^Q$, after some algebra, we find:

$$y_t^Q = y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-3}^M + y_{t-4}^M \tag{17.A.13}$$

which is the expression that tells us how monthly variables relates with quarterly variables.

Now, suppose that we have a vector of monthly growth rates $\mathbf{x}_t^M$ described by a factor model:

$$\mathbf{x}_t^M = \mathbf{\Lambda}_m \mathbf{F}_t + \boldsymbol{\xi}_t^M, \tag{17.A.14}$$

Suppose that we have a vector of quarterly growth rates $\mathbf{x}_t^Q$, the question is how can we construct a factor model for $\mathbf{x}_t^M$ and $\mathbf{x}_t^Q$? In other words, how it is possible to let coexist $\mathbf{x}_t^M$, $\mathbf{x}_t^Q$, $\mathbf{F}_t$ and $\boldsymbol{\xi}_t$? The answer is given in equation (17.A.13). Namely:

$$\begin{bmatrix} \mathbf{x}_t^M \\ \mathbf{x}_t^Q \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}_m & 0 & 0 & 0 & 0 \\ \mathbf{\Lambda}_q & 2\mathbf{\Lambda}_q & 3\mathbf{\Lambda}_q & 2\mathbf{\Lambda}_q & \mathbf{\Lambda}_q \end{bmatrix} \begin{bmatrix} \mathbf{F}_t \\ \vdots \\ \mathbf{f}_{t-4} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_t^M \\ \boldsymbol{\xi}_t^Q \\ \vdots \\ \boldsymbol{\xi}_{t-4}^Q \end{bmatrix}. \tag{17.A.15}$$

(17.A.15) is the state representation of the Dynamic Factor model with mixed frequency, which can be used to estimate the model with the Kalman Filter and the EM-algorithm.

17.A.5 Extensions of the Dynamic Factor Model

In this Section we will review two extensions of the Dynamic Factor model presented in Section 17.2, namely Bayesian Dynamic Factor models, and Non-Stationary Dynamic Factor models.

Bayesian Dynamic Factor Models

The Bayesian Dynamic Factor model (BDFM) is an extension to the Bayesian framework of the Dynamic Factor model presented in Section 17.2. Developed by D'Agostino et al. (2014), the BDFM was used by Luciani and Ricci (2014) for predicting Norwegian GDP. The BDFM is defined as follows:

$$x_{it} = \sum_{s=0}^{p} \boldsymbol{\lambda}_{is} \mathbf{F}_{t-s} + \sum_{s=1}^{p} \rho_{is} x_{it-s} + e_{it} \tag{17.A.16}$$

$$\mathbf{F}_t = \sum_{s=1}^{p} \mathbf{A}_s \mathbf{F}_{t-s} + \mathbf{u}_t \tag{17.A.17}$$

where $\mathbf{u}_t \sim \mathcal{N}(0, \mathbf{I}_r)$ and $e_{it} \sim \mathcal{N}(0, \psi_{it})$. The vector $\mathbf{x}_t$ includes both monthly and quarterly variables and this mixed frequency is accounted for as explained in Annex 17.A.4.

Compared with model (17.4)-(17.7), the intuition and the working of (17.A.16)-(17.A.17) is exactly the same, the only difference being that in (17.A.16) the factors are allowed to have an impact on the variables not only contemporaneously, but also dynamically through the polynomial $\boldsymbol{\lambda}_i(L) = \sum_{s=0}^{p} \boldsymbol{\lambda}_{is}$, and also that $p = 12$ much larger than what it is usually set for (17.4)-(17.7). This feature helps accommodating the dynamic heterogeneity of different variables, where the expression "dynamic heterogeneity of different variables" refers to the possibilities that some variables may be contemporaneously correlated with quarter-on-quarter GDP growth, while others may be correlated with different transformations of GDP. This is the case, for example, of survey indicators which tend to be more correlated with yearly GDP growth (i.e. the growth rate from a year ago) than with quarter-on-quarter GDP growth (see Figure 17.A.1).

Figure 17.A.1: Dynamic Heterogeneity: GDP vs PMI
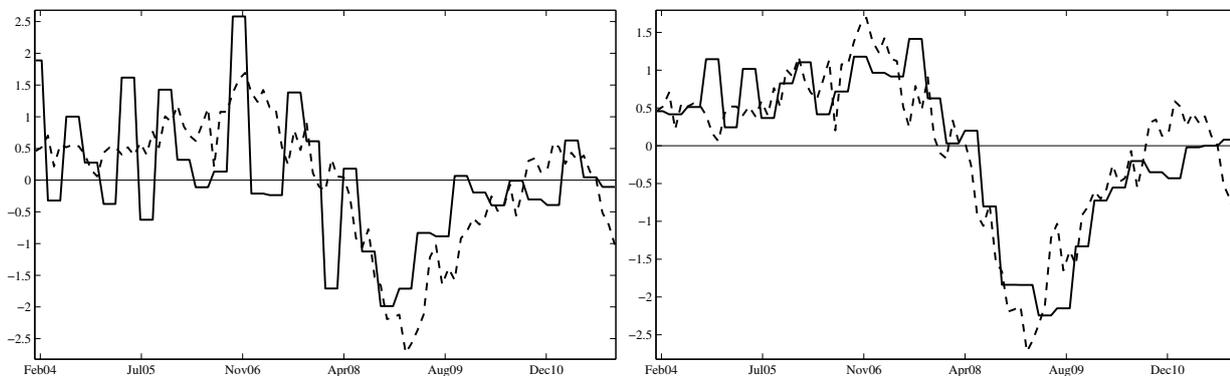
The left plot reports quarter-on-quarter Norwegian GDP growth and PMI, while the right plot reports yearly GDP growth and PMI. In both plots, the solid line is GDP, while the dashed line is PMI. Both variables are plotted at the monthly frequency so that the value of GDP growth is repeated three consecutive times.
Source: Luciani and Ricci (2014).

The use of Bayesian methods is motivated by the fact that model (17.A.16)-(17.A.17) requires to estimate a large number of parameters, which can easily lead to very volatile predictions due to estimation uncertainty, and Bayesian methods –by shrinking the factor model toward a simple naive prior model– are able to limit this uncertainty.[12]

Figure 17.A.2 reports the predictions of Norwegian GDP obtained by Luciani and Ricci (2014) with the BDFM.

---

[12]For example, Luciani and Ricci (2014) set the priors on the coefficients to have mean zero and such that the variance is smaller for higher order lags, so that posterior coefficients of high order lags are different from zero only if the data strongly point in that direction.

The model is estimated on 15 variables, where the variables were selected because are targeted by market operators (see Section 17.5.1). As we can see, compared to the other benchmark models the BDFM is the only model able to capture the downturn of 2009, and, moreover, the prediction of the BDFM is very close to that of the Bloomberg Surveys.

Figure 17.A.2: Predicting quarter-on-quarter GDP growth rate
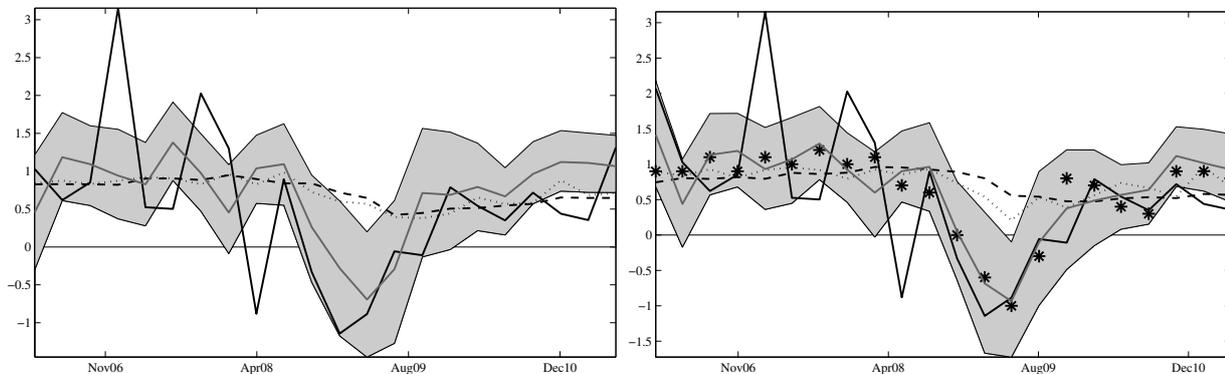
The left plot reports the prediction of Norwegian GDP growth conducted the last day of the current quarter, while the right plot report the prediction conducted the last day before GDP is released. In both plots, the circles are the actual GDP growth rate. The gray solid line is the point prediction obtained with the BDFM, while the shaded area is the 80% confidence band. The dashed line is the prediction obtained with the MIDAS model, while the dotted line is the prediction obtained with the Bridge model. Finally, in the right plot the asterisks are the prediction from the Bloomberg Survey.
Source: Luciani and Ricci (2014).

Bayesian methods for estimating Dynamic Factor models have been used also by Marcellino et al. (2013) who estimate a model similar to (17.6)-(17.8) in which both the factors and the idiosyncratic components have stochastic volatility. However, compared to D'Agostino et al. (2014) who use Bayesian methods to reduce the estimation uncertainty brought about the over-parametrization of the model, Marcellino et al. (2013) use Bayesian methods to introduce stochastic volatilities.

Marcellino et al. (2013) estimate their model on a small database of nine variables which were pre-selected from a large database of Euro Area variables using an algorithm similar to that of Camacho and Perez-Quiros (2010) described in 17.5.1. The model is used to predict Euro Area GDP and results show that the introduction of stochastic volatility significantly contributes to an improvement in density forecast accuracy.

In summary, the evidence available in the literature indicates that Bayesian Dynamic Factor models perform well in forecasting. To the best of our knowledge, though, in the literature there is no formal comparison between Bayesian and *Frequentist* Dynamic Factor models, so that it is very difficult to asses whether Bayesian techniques do really perform well in medium-size information sets, or if the gain in forecasting is negligible. However, despite in terms of forecasting performance it is impossible to draw a conclusion, it is clear that Bayesian Dynamic Factor models are more flexible as they, for example, allow for dynamic factor loadings, or to account for stochastic volatility. Therefore, the choice of using a Bayesian or a *Frequentist* Dynamic Factor model seems to be very case specific since it depends on the data at hand, or/and on the kind of features that the researcher wants to capture.

Non-Stationary Dynamic Factor Models

All the literature that we reviewed so far is based on stationary data. Indeed, with very few exceptions Bai (2004); Bai and Ng (2004); Peña and Poncela (2004b), the factor literature has studied only the stationary case. However, recently Barigozzi et al. (2014a) generalized model (17.4)-(17.7) to the non-stationary case,

and Barigozzi et al. (2014b) studied estimation of the factors $\mathbf{F}_t$ when they are $I(1)$ and when $\boldsymbol{\xi}_t$ may or may not be $I(0)$. Formally, let $\mathbf{x}_t$ be a vector of $I(1)$ variables, then the non-stationary Dynamic Factor model is defined as follows:

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t \tag{17.A.18}$$

$$\boldsymbol{\chi}_t = \boldsymbol{\Lambda}\mathbf{F}_t \tag{17.A.19}$$

$$\mathbf{F}_t = \mathbf{F}_{t-1} + \mathbf{B}(L)\mathbf{u}_t \tag{17.A.20}$$

$$\boldsymbol{\xi}_t = \boldsymbol{\xi}_{t-1} + \mathbf{D}(L)\mathbf{e}_t \tag{17.A.21}$$

where the notation and dimensions are the same as for model (17.4)-(17.7).

Barigozzi et al. (2014a) prove under very mild conditions that when $q < r$, generically $\Delta\mathbf{F}_t$ admits a finite autoregressive representation with $r - q + d$ error corrections, where d is the number of common shocks that have no-permanent effects on $\mathbf{x}_t$. However, they use this result for estimating impulse response functions, while not considering forecasting.

Model (17.A.18)-(17.A.21) so far has not been used for forecasting. There are two exceptions to this statement. The first one is Banerjee et al. (2013), who, however, use a restricted version of (17.A.18)-(17.A.21) where $\boldsymbol{\xi}_t \sim I(0)$, which implies that all the variables in the panel are co-integrated. The second one is Peña and Poncela (2004a), who, however, consider a non-stationary exact factor model –with $\boldsymbol{\xi}_t \sim I(0)$– estimated on a small database. Notwithstanding the limitations of these analyses, the forecasting performance of these models are good thus indicating that the study of the non-stationary Dynamic Factor model in forecasting is a promising line of research.

Bibliography

Aastveit K.A., Trovik T. (2012), Nowcasting norwegian GDP: the role of asset prices in a small open economy. Empirical Economics 42 (1), 95–119.

Altissimo F., Bassanetti A., Cristadoro R., Forni M., Hallin M., Lippi M., Reichlin L., Veronese G. (2001), A real time coincident indicator for the euro area business cycle. Discussion Paper 3108, CEPR.

Altissimo F., Cristadoro R., Forni M., Lippi M., Veronese G. (2010), New Eurocoin: Tracking economic growth in real time. The Review of Economics and Statistics 92 (4), 1024–1034.

Anderson B. D., Deistler M. (2008), Generalized linear dynamic factor model–a structure theory. IEE Conference on Decision and Control.

Anderson B. D., Deistler M. (2008), Properties of zero-free transfer function matrices. SICE Journal of Control, Measurement and System Integration 1, 284–292.

Angelini E., Bańbura M., Rünstler G. (2010). Estimating and forecasting the Euro Area monthly national accounts from a dynamic factor model. OECD Journal: Journal of Business Cycle Measurement and Analysis 2010 (1), 1–22.

Angelini E., Camba-Mendez G., Giannone D., Reichlin L., Rünstler G. (2011), Short-term forecasts of euro area gdp growth. The Econometrics Journal 14 (1), C25–C44.

Artis M.J., Banerjee A., Marcellino M. (2005), Factor forecasts for the UK. Journal of Forecasting 24 (4), 279–298.

Bai J. (2003), Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai J. (2004), Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics 122, 137–183.

Bai J., Ng S. (2002), Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai J., Ng S. (2004), A PANIC attack on unit roots and cointegration. Econometrica 72, 1127–1177.

Bai J., Ng S. (2006), Confidence intervals for diffusion index forecasts and inference for Factor-Augmented Regressions. Econometrica 74, 1133–1150.

Bai J., Ng S. (2008a), Forecasting economic time series using targeted predictors. Journal of Econometrics 146 (2), 304–317.

Bai, J. and S. Ng (2008b), Large dimensional factor analysis. Foundations and Trends in Econometrics 3 (2), 89–163.

Bańbura M., Giannone D., Modugno M., Reichlin L. (2012), Now-casting and the real-time data-flow. In G. Elliott and A. Timmermann (Eds.), Handbook of Economic Forecasting. Amsterdam: Elsevier-North Holland, 2012.

Bańbura M., Giannone D., Reichlin L. (2010), Large Bayesian vector auto regressions. Journal of Applied Econometrics 25 (1), 71–92.

Bańbura M., Giannone D., Reichlin L. (2011), Nowcasting. In M. P. Clements and D. F. Hendry (Eds.), Oxford Handbook on Economic Forecasting. New York: Oxford University Press.

Bańbura M., Modugno M. (2014), Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. Journal of Applied Econometrics 29 (1), 133–160.

Bańbura M., Rünstler G. (2011), A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. International Journal of Forecasting 27 (2), 333–346.

Banerjee A., Marcellino M., Masten I. (2013), Forecasting with factor augmented error correction models. International Journal of Forecasting, 2013.

Barhoumi K., Darné O., Ferrara L. (2010), Are disaggregate data useful for factor analysis in forecasting French GDP? Journal of Forecasting 29 (1-2), 132–144.

Barigozzi M., Conti A. M., Luciani M. (2014), Do Euro Area countries respond asymmetrically to the common monetary policy? Oxford Bulletin of Economics and Statistics 76 (5), 693–714.

Barigozzi M., Lippi M., Luciani M. (2014a), Dynamic factor models, cointegration, and error correction mechanisms. Working Papers ECARES 2014-14, Université Libre de Bruxelles.

Barigozzi M., Lippi M., Luciani M. (2014b), Estimating I(1) dynamic factor models. Université libre de Bruxelles.

Baxter M., King R.G. (1999), Measuring business cycles: Approximate bandpass filters for economic time series. The Review of Economics and Statistics 81 (4), 575–593.

Bernanke B.S., Boivin J. (2003), Monetary policy in a data-rich environment. Journal of Monetary Economics 50, 525–546.

Boivin J., Ng S. (2005), Understanding and Comparing Factor-Based Forecasts. International Journal of Central Banking 1 (3), 117–151.

Boivin J., Ng S. (2006), Are more data always better for factor analysis? Journal of Econometrics 127 (1), 169–194.

Brave S., Butters R.A. (2014), Nowcasting using the Chicago Fed National Activity Index. Economic Perspectives 38 (1), 19–37. Federal Reserve Bank of Chicago.

Brave S., Lichtenstein M. (2012), A different way to review the Chicago Fed National Activity Index. Chicago Fed Letter 298, Federal Reserve Bank of Chicago.

Breitung J., Tenhofen J. (2011), GLS Estimation of Dynamic Factor Models. Journal of the American Statistical Association 106, 1150-.1166

Camacho M., Perez-Quiros G. (2010), Introducing the euro-sting: Short-term indicator of euro area growth. Journal of Applied Econometrics 25 (4), 663–694.

Choi I. (2012), Efficient Estimation of Factor Models. Econometric Theory 28, 274–308.

Corielli F., Marcellino M. (2006), Factor based index tracking. Journal of Banking & Finance 30, 2215–2233.

Cristadoro R., Forni M., Reichlin L., Veronese G. (2005), A core inflation indicator for the Euro Area. Journal of Money Credit and Banking 37 (3), 539–560.

D'Agostino A., Giannone D. (2012), Comparing alternative predictors based on large-panel factor models. Oxford Bulletin of Economics and Statistics 74 (2), 306–326.

D'Agostino A., Giannone D., Lenza M., Modugno M. (2014), Nowcasting Business Cycles: A Bayesian Approach to Dynamic Heterogenous Factor Models. Université libre de Bruxelles.

D'Agostino A., McQuinn K., O'Brien D. (2012), Now-casting Irish GDP. OECD Journal: Journal of Business Cycle Measurement and Analysis 2012 (2), 21–31.

Dempster A., Laird N., Rubin D. (1977), Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B 39 (1), 1–38.

Doz C., Giannone D., Reichlin L. (2011), A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics 164 (1), 188–205.

Doz C., Giannone D., Reichlin L. (2012), A quasi maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics 94 (4), 1014–1024.

Engle R.F., Watson M. W. (1981), A one-factor multivariate time series model of metropolitan wage rates. Journal of the American Statistical Association 76 (376), 774-781.

Evans C.L., Liu C.T., Pham-Kanter G. (2002), The 2001 recession and the Chicago Fed National Activity Index: Identifying business cycle turning points. Economic Perspectives 26 (3), 26–43. Federal Reserve Bank of Chicago.

Forni M. and L. Gambetti (2010), The dynamic effects of monetary policy: A structural factor model approach. Journal of Monetary Economics 57, 203–216.

Forni M., Giannone D., Lippi M., Reichlin L. (2009), Opening the Black Box: Structural Factor Models versus Structural VARs. Econometric Theory 25 (5), 1319–1347.

Forni M., Hallin M., Lippi M., Reichlin L. (2000), The Generalized Dynamic Factor Model: Identification and Estimation. The Review of Economics and Statistics 82, 540–554.

Forni M., Hallin M., Lippi M., Reichlin L. (2003), Do financial variables help forecasting inflation and real activity in the Euro Area? Journal of Monetary Economics 50 (6), 1243–1255.

Forni M., Hallin M., Lippi M., Reichlin L. (2004), The Generalized Dynamic Factor Model: Consistency and rates. Journal of Econometrics 119, 231–255.

Forni M., M. Hallin, M. Lippi, and L. Reichlin (2005). The Generalized Dynamic Factor Model: One Sided Estimation and Forecasting. Journal of the American Statistical Association 100 (4), 830–840.

Forni M., Hallin M., Lippi M., Zaffaroni P. (2014), Dynamic factor models with infinite- dimensional factor spaces: one-sided representations. Journal of Econometrics, 2014.

Forni M., Lippi M. (2001), The Generalized Dynamic Factor Model: Representation theory. Econometric Theory 17, 1113–1141.

Forni M., Lippi M. (2010), The unrestricted dynamic factor model: One-sided representation results. Journal of Econometrics 163, 23–28.

Geweke J. (1977), The dynamic factor analysis of economic time series. In D. J. Aigner and A. S. Goldberger (Eds.), Latent Variables in Socio-Economic Models. Amsterdam: North Holland.

Giannone D., Matheson T. D. (2007), A new core inflation indicator for new zealand. International Journal of Central Banking 3 (4), 145–180.

Giannone D., Reichlin L., Sala L. (2005), Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), NBER Macroeconomics Annual 2004. MIT Press.

Giannone D., Reichlin L., Small D. (2008), Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics 55 (4), 665–676.

Hallin M., Mathias C., Pirotte H., Veredas D. (2011), Market liquidity as dynamic factors. Journal of Econometrics 163, 42–50.

Lawley D.N., Maxwell A.E. (1971), Factor analysis as a statistical method (2nd ed.). London: Butterworths.

Luciani M. (2015), Monetary policy and the housing market: A structural factor analysis. Journal of Applied Econometrics 30 (2), 199–218.

Luciani M. (2014), Forecasting with approximate dynamic factor models: the role of non-pervasive shocks. International Journal of Forecasting 30 (1), 20–29.

Luciani M., Ricci L. (2014), Nowcasting Norway. International Journal of Central Banking 10 (4), 215–248.

Luciani M., Veredas D. (2015), Estimating and forecasting large panels of volatilities with approximate dynamic factor models. Journal of Forecasting 34 (3), 163–176.

Ludvigson S.C., Ng S. (2007), The empirical risk-return relation: a factor analysis approach. Journal of Financial Economics 83, 171–222.

Ludvigson S.C., Ng S. (2009), Macro factors in bond risk premia. The Review of Financial Studies 22, 5027–5067.

Marcellino, M., M. Porqueddu, and F. Venditti (2013). Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility. Temi di Discussione 896, Banca D'Italia.

Marcellino M., Schumacher C. (2010), Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. Oxford Bulletin of Economics and Statistics 72 (4), 518–550.

Marcellino M., Stock J. H., Watson M. W. (2003), Macroeconomic forecasting in the Euro Area: Country specific versus area-wide information. European Economic Review 47 (1), 1–18.

Mariano R.S., Murasawa Y. (2003), A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics 18 (4), 427–443.

Matheson T. D. (2010). An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys. Economic Modelling 27 (1), 304–314.

Modugno, M. (2014). Now-casting inflation using high frequency data. International Journal of Forecasting 29 (4), 664–675.

Onatski A. (2012), Asymptotics of the principal components estimator of large factor models with weakly influential factors. Journal of Econometrics 168, 244–258.

Peña D., Poncela P. (2004), Forecasting with non stationary dynamic factor models. Journal of Econometrics 119, 291–321.

Peña D., Poncela P. (2004), Nonstationary dynamic factor analysis. Journal of Statistical Planning and Inference 136, 1237–1257.

Rünstler G., Barhoumi K., Benk S., Cristadoro R., Den Reijer A., Jakaitiene A., Jelonek P., Rua A., Ruth K., Van Nieuwenhuyze C. (2009), Short-term forecasting of GDP using large datasets: a pseudo real-time forecast evaluation exercise. Journal of Forecasting 28 (7), 595–611.

Sargent T.J., Sims C.A. (1977), Business cycle modeling without pretending to have too much a-priori economic theory. In C. A. Sims (Ed.), New Methods in Business Cycle Research. Minneaopolis: Federal Reserve Bank of Minneapolis.

Schumacher C. (2007), Forecasting German GDP using alternative factor models based on large datasets. Journal of Forecasting 26 (4), 271–302.

Schumacher C. (2010), Factor forecasting using international targeted predictors: The case of German GDP. Economics Letters 107, 95–98.

Schumacher C., Breitung J. (2008), Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. International Journal of Forecasting 24 (3), 386–398.

Stock J.H., Watson M.W. (1989), New indexes of coincident and leading economic indicators. In O. J. Blanchard and S. Fischer (Eds.), NBER Macroeconomics Annual 1989. MIT Press.

Stock J.H., Watson M.W. (1999), Forecasting inflation. Journal of Monetary Economics 44, 293–335.

Stock J.H., Watson M.W. (2002), Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock J.H., Watson M.W. (2002), Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics 20 (2), 147–162.

Stock J.H., Watson M.W. (2005), Implications of dynamic factor models for VAR analysis. Working Paper 11467, NBER.

Stock J.H., Watson M.W. (2011), Dynamic factor models. In M. P. Clements and D. F. Hendry (Eds.), Oxford Handbook of Forecasting. Oxford University Press.

Watson M.W., Engle R.F. (1983), Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. Journal of Econometrics 23, 385–400.

18

Regression, Principal Components and Small Scale Factors Based Models

Author

Françoise Charpin
L'Institut d'Études Politiques de Paris (Sciences Po)
françoise.charpin@sciences-po.fr

Table of Contents

18.1 Introduction

The methods presented in this chapter are designed to give an early estimate of macroeconomic aggregates using monthly measures of real economic activity. Here, they will be applied to quarterly growth rate of GDP that to be estimated for the coincident quarter. But these methods can also be employed to forecast the GDP growth of the next quarter. The very short-term forecasting horizon comes from the fact that models used here are not behavioural. They just rely on empirical links between short-term indicators and a quarterly national account aggregate. Short-term indicators are monthly data available before the release of the aggregate to estimate. This monthly information comprises survey data (soft data), monthly macroeconomic data (hard data) and financial data, more generally all high frequency series that are likely to bring information about current and future activity. National institutes release a flash (or a first) estimate of GDP for quarter T several days after the end of quarter T, say H days, at a date we denote $(T + H)$ in the following. For example, the first estimate of the US GDP of quarter T is released at $(T + 30)$ and the flash estimate of the euro area GDP of quarter T, at $(T + 45)$. Early estimates of quarter T will occur between dates $(T - 1 + H)$ and $(T + H)$. They can be provided for example at the end of each week over the period $[(T - 1 + H), (T + H)]$ or at the end of each month. They may also be accompanied by a forecast of the next quarter.

Survey data and financial data are particularly interesting because they are rapidly available and hardly or not at all revised. But estimates based exclusively on these two data types are generally not fully accurate. Moreover some hard data (as industrial production index) are directly used by quarterly national accounts to elaborate their estimate or present a good correlation with their national accounts counterpart (as monthly exports and imports). Thus it is logical to include hard data in the estimation process although the latter are subject to revisions, are very volatile and thus very difficult to forecast. Nevertheless the hard data monthly growth rate displays generally a strong autoregressivity, a characteristic that can be exploited in the very short term, for extrapolating them one or two months ahead.

In this chapter, the early estimate of the coincident quarter is carried out using one GDP equation completed by auxiliary equations to generate forecasts of missing monthly observations on the quarter to estimate. These equations are regressions with either specific series or factors as regressor. These factors are principal components extracted from a small dataset, composed of series selected by using the algorithm of the Least Angle Regression.[1] Factors are used in order to remedy the collinearity between monthly data, especially between survey data. Thus, it allows the prediction to be based on wider information than when individual series are employed as regressor. When the GDP growth rate is regressed on factors, the estimate of the next quarter can be obtained by extrapolating the factors with $AR(p)$ models or with a VAR.

In this chapter, the spirit is different from that of large scale factor models (see chapter 17). The first difference concerns the size of the dataset from which the factors are extracted. Several papers (Boivin and Ng (2006), Bai and Ng (2008)) mention that the introduction of many series, more or less related to the target (here GDP growth), can produce a noise which deteriorates the estimates. That is why, following Bai and Ng, we propose to consider only the series directly related to the target[2], approximately all the series which can help to predict the target but which cannot be introduced simultaneously in a regression because of multi-collinearity. The second difference concerns the treatment of missing monthly observations on the quarter to estimate. In some large scale factor model, no auxiliary equations are employed to handle the problem of missing observations which is solved internally i.e. the estimation method allows you to compute together forecast of these monthly series in addition to that of the GDP growth. This is obviously very convenient but the same criticism as that mentioned previously can be addressed. The forecast of monthly data is then based on many series more or less related to the series to forecast and, more serious, the strong autoregressivity of these monthly series within a quarter is not taken into account whereas this is an important feature of these series. Forgetting it should affect the accuracy of forecasts.

---

[1] Least Angle Regression, Efron et al. (2004), shortly described in the annex.
[2] Called by the authors the targeted series.

Note also that the models employed in this chapter are not of MIDAS types. [3] These latter allow the regress and regressors to be sampled at different frequencies and they rely solely on released monthly information. Thus the conversion of monthly to quarterly does not arise as well as the forecast of missing monthly data. Despite these conveniences, the superiority of these models is not obvious.

The chapter is organized as follows. Section 18.2 surveys some data issues (frequency conversion, stationarity and exogeneity). Section 18.3 is devoted to GDP equations using specific series. These equations (or bridge models) may or may not exploit cointegrating relations between the target (in logarithm) and some hard data (in logarithm). In other words, two types of bridge models can be found, those which include error correction terms and others without. Section 18.4 deals with principal components regressions and the selection of targeted series entering factors. Section 18.5 describes the forecast of missing monthly observations. Section 18.6 surveys MIDAS models. Conclusions are provided in section 18.7 at the end, alongside an annex in section 18.A.

18.2 Data issues

We are considering in turn the three types of data (hard data, soft data and financial data), each having its own features. Stationarity and exogeneity are needed when estimating the GDP equation. The conversion problem arises because the GDP is quarterly and the regressors are monthly.

18.2.1 About hard data

When the target is the GDP, the monthly macroeconomic data of interest are those that present a good correlation with the aggregates of the accounting identity (1) giving GDP. It is logical to try to do as national institutes to provide an early GDP estimate.

$$GDP = \text{private consumption} + \text{public consumption} + \text{investment} + \text{inventory changes}$$
$$+ \text{exports} - \text{imports} \qquad (18.1)$$

For some demand components of (1), like exports and imports, information is available on a monthly basis. In some countries, it is also the case of private consumption, otherwise retail sales can be used as a proxy. In addition, on the supply side, industrial production is known on a monthly basis.

The question of frequency conversion does not arise with this type of data because in each case, a logical conversion exists. For example, the quarterly consumption is obtained by summing monthly consumption over the quarter. Thus the temporal aggregation of hard data is done by averaging in case of index (IPI for example) or summing in case of flow.

All these hard data are $I(1)$ and must be considered in growth rate except if cointegrating relations with GDP can be detected so that the log levels can be kept. Concerning exogeneity, as these data are released before GDP and as some will be used to construct the GDP first estimate, they can be considered as exogenous.

18.2.2 About survey data

Most of them are $I(0)$ and can be used directly without transformation. Nevertheless, it is common to find that the change in the agent's opinion is more significant for prediction that the opinion itself. Thus, even if the series is $I(0)$, its first difference may be more interesting to predict future. All can be considered as exogenous.

---

[3]Mixed-data sampling models, introduced by Ghysels et al. (2005), Ghysels et al. (2006).

At the end of month t (or at the beginning of month $(t + 1)$ for US survey), survey data affected to month t are published. But the corresponding surveys are conducted during the first 2 or 3 weeks of month t. Thus, the survey questionnaire may very well be filled during the first week of month t. Then, when asked to assess the current business situation, the response concerns at best the situation at the end of the previous month. Thus, a "current" assessment could in fact better characterize month (t - 1) as month t and a "past 3 months" assessment, months $(t - 3)$, $(t - 2)$, and $(t - 1)$. If, now, we take the question about a "future" assessment (as production the next 3 months), the answer should cover months $(t + 1)$, $(t + 2)$ and $(t + 3)$, especially in the case where the questionnaire is filled the third week of month t. It explains why question about a "future" assessment are often found as coincident instead of leading and question about a "current" assessment, as lagging instead of coincident.

The conversion problem has no more a natural solution as previously. It can be treated in different ways. These are (i) averaging the months of the quarter (ii) proceed to a lead of the series before averaging (the first quarter, for example, is then represented by the survey data released in February, March and April) given that the survey questionnaire is filled during the month t as explained above (iii) taking the latest known value to better anticipate future opinion. To choose between these options, empirical tests can be carried out. Our own experience leads us to adopt the first two ways (i.e. averaging) to reduce the volatility of monthly time series. Note that it is possible to avoid the conversion problem by adopting a model for mixed-frequency data (Section 18.6).

18.2.3 About financial data

These are: interest rates, interest rates spread, stock index, exchange rates and oil price. These hard data are $I(1)$ except interest rate spreads which are $I(0)$. Stock index, exchange rates and oil price, all being computed in real terms, must be considered in growth rate, while interest rates, in first difference. Exchange rates and oil price can be declared exogenous. The issue of exogeneity therefore arises for interest rates, interest rates spread, and stock index. However, these variables are introduced with several lags in GDP equation and this predetermination avoids the issue of exogeneity.

The quarterly frequency conversion by averaging financial data over the quarter has one advantage that of reducing the growth rate volatility of these series. For this reason, this is what we advocate. But, of course, one alternative is to adopt a MIDAS model.[4]

18.3 Bridge equations

Our methodology is described in Section 18.3.1. It is illustrated by two examples. One, in Section 18.3.2, concerns the French GDP and, the other, in Section 18.3.3, the US GDP.

18.3.1 Description of bridge equations

Bridge equations are regressions that link high frequency indicators (monthly hard and soft data) to low frequency variables (here quarterly GDP growth). It allows computing early estimates of GDP as monthly indicators are released before GDP. Since the monthly indicators are usually only partially available over the prediction period, the estimations of quarterly GDP growth are obtained in two steps. First, monthly indicators are forecasted over the remainder of the quarter (Section 5), and then aggregated to obtain their quarterly values. Second, the aggregated values are used as regressors in the bridge equation in order to obtain estimates of GDP growth. There are many studies of this type in the US and European countries conducted

---

[4]The level of financial data.

by Central Banks or institutes of economic forecasts. It is commonly recognized by practitioners that these simple models give results at least as good as the more sophisticated models of Chapter 17.

The questions that arise when developing such models are (i) the presence or not of hard data among the monthly indicators (ii) the conversion or not of monthly series to quarterly (iii) the selection of monthly indicators to include in the bridge model (iv) the presence of cointegrating relations between GDP and hard data.

On point (i) we have given our opinion in the introduction, namely, do as national institutes when preparing their first (or flash) estimates, i.e. use hard data. Point (ii) is discussed in Sections 2 and 6. Point (iii), the selection of monthly indicators to include in the bridge model, is usually based on a general-to-specific algorithm and relies on different econometric criteria but also on real time or pseudo real time performances of the model.

In relation to point (iii), the LARS algorithm can also be used as explained now. A large dataset is formed containing all monthly indicators, each with several lags and possibly several forms for survey data. These are: level z, first difference Δz and non linear first difference $\Delta z \times |\Delta z|$. This latter transformation makes the regression coefficient of Δz to depend on the absolute change in z. This can help to explain the volatility of growth and also the amplitude of the last recession. The first twenty ranked series are retained. The selection of monthly indicators to include in the bridge model is done from this small dataset.[5]

We discuss now point (iv). As monthly hard data representing some components of the supply side or the demand side are available, equilibrium relationships may exist between GDP and some monthly proxies of demand components or supply components. One can seek to exploit these cointegrating relations. If one (or more) cointegrating relation is found, an error correction bridge model can be estimated.

For example, consider these four monthly components available in USA and France: Industrial production index, real private consumption, real exports, real imports.[6] If the Johansen test of cointegration (Johansen (1995)) is run on the five components in logarithm, GDP plus the four previous series in quarterly terms, one cointegrating relation is detected between them. Thus an error correction term can be added to the bridge model. If two cointegrating relations were found, the inclusion of two error correction terms has to be tested.

It is not always easy to determine the number of cointegrating relations. The Johansen's test turns out to be very sensitive to the slightest modification of the context: order of VAR, choice of the deterministic part, Bartlett's correction of the trace test or no correction. Moreover, the test is not adapted to the present situation where we have only one endogenous variable (GDP) all others being considered as exogenous. By cons, it could be adapted to determine the number of cointegrating relations between the four hard data: industrial production index, real consumption, real exports, real imports. It is interesting to explore this topic because when you find more than one cointegrating relation within the VAR of 5 components, one or several of them do not necessarily contain the GDP.

Given our specific context, only one endogenous variable, a univariate test of cointegration is appropriate but does not provide information on the number of cointegrating relations between the five components. That is why one can first try to determine this number. Then, we advise to follow the approach described below. If y denotes the GDP in log, xh the h hard data (quarterly converted) in log and z_k some I(0) variable, the following error correction model is estimated (equation 18.1):

$$\Delta y = c + \alpha y_{-1} + \sum_h a_h x_{h,-1} + \beta \Delta y_{-1} + \sum_h b_{h0} \Delta x_h + \sum_h b_{h1} \Delta x_{h,-1} + \sum_k \sum_{s=0}^{S} d_{ks} z_{k,-s} \qquad (18.1)$$

In equation 18.1 we limit the lags of Δy and Δx_h to one because it is appropriate for our context. The cointegration test is carried out with the T-ratio of coefficient α and the Ericsson and MacKinnon (2002). When cointegration is not accepted, the usual bridge model (equation 18.2) prevails:

---

[5]This is an order of magnitude, in general the significant regressors are there.
[6]We choose these two countries because they are the only ones where a monthly personal consumption is available.

$$\Delta y = c + \beta \Delta y_{-1} + \sum_h b_{h0} \Delta x_h + \sum_h b_{h1} \Delta x_{h,-1} + \sum_k \sum_{s=0}^{S} d_{ks} z_{k,-s} \tag{18.2}$$

If we know a priori that there is only one cointegrating relation, it is given by equation 18.1. In cases of several cointegrating relations and at least two significant correction terms, then equation 18.1 will give a linear combination of cointegrating relations involved. This is not annoying to the extent that the GDP equation is not behavioural and serves solely to make forecasts.

18.3.2 Estimation of the French GDP growth by a bridge error correction model

The quarterly data used cover the period 1989-2012.

First step: Number of cointegrating relation in the 5-dimensional VAR model

For this investigation, we choose a VAR(3) with unrestricted constants in the equations. Results of the rank test are given in figure 18.1. They are obtained with Cats 2.0 in WinRats 8.0.

Figure 18.1: Rank test in a 5-dimensional VAR(3)

```
VARIABLE NAME
LGDP GDP in logarithm
LIPI IPI in logarithm
LIMPOR imports in logarithm
LEXPOR exports in logarithm
LCONS  consumption in logarithm

MODEL SUMMARY
Sample:                         1989:01 to 2012:01 (93 observations)
Effective Sample:               1989:04 to 2012:01 (90 observations)
Obs. - No. of variables:        74
System variables:               LGDP LIPI LIMPOR LEXPOR LCONS
Constant/Trend:                 Unrestricted Constant
Lags in VAR:                    3

I(1)-ANALYSIS
 p-r r Eig.Value Trace  Trace* Frac95 P-Value P-Value*
   5  0    0.411 95.179 85.101 69.611   0.000    0.001
   4  1    0.209 47.051 42.054 47.707   0.058    0.158
   3  2    0.157 25.722 22.835 29.804   0.141    0.263
   2  3    0.071 10.216  7.815 15.408   0.269    0.493
   1  4    0.038  3.484  2.802  3.841   0.062    0.094

Trace*, P-Value*: small sample correction (Doornik, 1998)

THE MATRICES BASED ON 1 COINTEGRATING VECTOR:
         LPIB     LIPI    LIMPOR    LEXPOR     LCONS
Beta(1) 1.000   -0.048    0.239    -0.363    -0.462
        (.NA)  (-2.500)  (5.726) (-13.708) (-13.597)
```

The test for $r = 0$ cointegrating relation is clearly rejected. The test for $r = 1$ is accepted. The cointegrating relation is given at the bottom of figure 18.1.

To confirm that, we seek for the number of cointegrating relation in the 4-dimensional VAR model formed with the four hard data components. For that, we choose a VAR(3) with unrestricted constants as previously. Results of the rank test are provided in figure 18.2. The test for $r = 0$ is accepted.

Second step: Estimation of a bridge error correction model

The bridge ECM is given by equation 18.2 Section 18.3.1. Its estimation is provided in Table 18.1. Variables z_k are sought among survey data and financial data. Those that are retained are listed in Table 18.1 with

Figure 18.2: Rank test in a 4-dimensional VAR(3)

```
Sample:                      1989:01 to 2012:02 (94 observations)
Effective Sample:            1989:04 to 2012:02 (91 observations)
Obs. - No. of variables:     78
System variables:            LIPI LIMPOR LEXPOR LCONS
Constant/Trend:              Unrestricted Constant
Lags in VAR:                 3

I(1)-ANALYSIS
 p-r r Eig.Value Trace  Trace* Frac95 P-Value P-Value*
  4  0    0.175 38.019 34.223 47.707  0.305    0.494
  3  1    0.133 20.554 18.305 29.804  0.397    0.554
  2  2    0.065  7.563  5.799 15.408  0.520    0.721
  1  3    0.016  1.451  1.222  3.841  0.228    0.269
```

their transformation and their lag. Three hard data enter the equation (18.2), consumption, exports and industrial production (Table 18.1). The imports are excluded from the equation. This was not the case of the cointegrating relation of the VAR(3) as shown in Table 18.1. As mentioned in section 18.3, what prevails for us is results obtained in the univariate framework with all determinants especially the variables z_k.

The T-Stat of the coefficient of the GDP in log indicates that there exists a cointegrating relation between the four I(1) variables (the critical value for size 5% is equal to -3.8). The other regressors are three survey data and one financial variable. The adjustment is very accurate as shown in Figure 1 ($\bar{R}^2 = 0.90$, SEE=0.16%). With this equation, the main source of errors is from prediction of the monthly indicators. With real time data, the Root Mean Square Error (RMSE) is greater than 0.16% and lies between 0.2 and 0.25% according to the estimation date.

If we examine the stability of the equation, we find a break in 2009Q3. When the estimation period ends in 2009Q2 (or before), the IPI in log does not belong to the co-integrating relation.[7] Whereas if the estimation period ends in 2009Q3 or later, the IPI in log belongs to the co-integrating relation. This unstability (inherent at practically all equations and due to the exceptional amplitude of the financial crisis is not necessary dramatic as the GDP equation is re-estimated before each estimation, which allows us to modify it in case of problems.

Figure 18.3: The French quarterly GDP growth rate observed and fitted (dotted line)
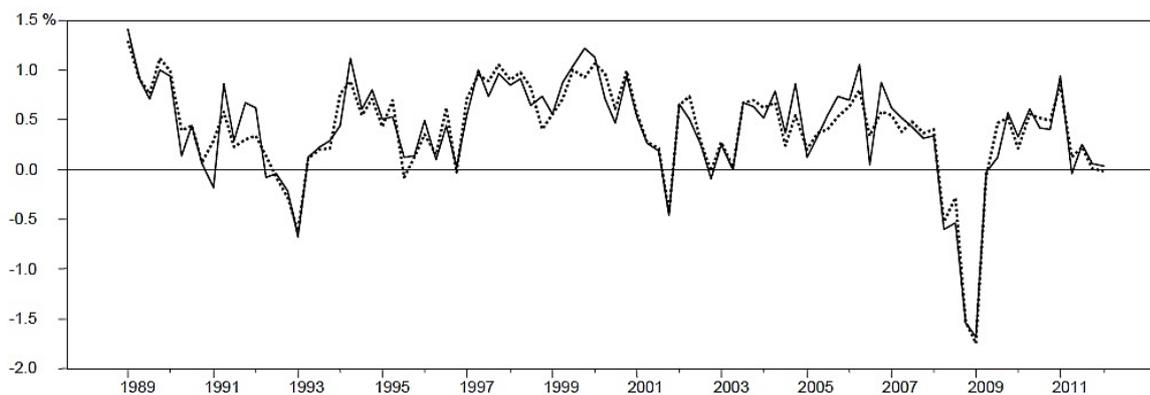
18.3.3 Estimation of the US GDP growth by a bridge error correction model

The quarterly data used cover the period 1990-2012.

---

[7]It has a negligible coefficient equal to 0.006 with a T-stat of 0.8.

Table 18.1: Estimation of a bridge error correction model for the French GDP

Variables	Transformation	Lag in quarter	Coefficient	T-Stat(*)
GDP (INSEE)	Log. level	1	-0.183	-5.7
Consumption (engineered products) (IN-SEE)	Log. level	1	0.050	5.5
Real exports (INSEE for exports, OFCE for real exports)	Log. level	1	0.037	4.9
Industrial production index (INSEE)	Log. level	1	0.012	2.5
Consumption (engineered products) (IN-SEE)	Log. diff.	0	0.077	6.7
Real exports (INSEE for exports, OFCE for real exports)	Log. diff.	0	0.036	3.4
Industrial production index (INSEE)	Log. diff.	0	0.113	6.1
Manufacturing output, next 3 months (Industry survey, INSEE)	First difference	0	0.116×10^3	3.4
Major purchases intentions, next 12 months (Consumer survey, INSEE)	Non linear first difference	2	0.034×10^3	3.1
Confidence indicators (Construction survey, DGECFIN)	First difference	2	0.084×10^3	2.3
French real stock index (MSCI)	Log. diff.	1	0.081×10^3	3.3
Constant			1.620	5.5

(*) The T-Stat of I(1) variables follow a non standard distribution. A table is available for the coefficient α in equation (2).

First step: Number of cointegrating relation in the 5-dimensional VAR model

For this investigation, we choose a VAR(4) with unrestricted constants in the equations. Results of the rank test are given in figure 18.4.

The trace test leads to accept $r = 1$ whereas the small sample corrected trace test leads to $r = 0$. Nevertheless, if we choose $r = 1$, we see that LIPI can be excluded from the cointegrating relation. Finally, the cointegrating relation (at the bottom of Table 18.4) contains two major variables LCONS and LIMPOR and a minor one, LEXPOR, less significant.

As previously, we seek for the number of cointegrating relation in the 4-dimensional VAR model formed with the four hard data components. For that, we choose a VAR(4) with unrestricted constants as previously. Results of the rank test are provided in figure 18.5. The hypothesis $r = 0$ is accepted.

Second step: Estimation of a bridge error correction model

The estimation of the bridge ECM is provided in figure 18.4. Two hard data enter the cointegrating relation, consumption and imports. In the VAR(4) model of figure 18.4 only one variable was excluded (industrial production index). The exports were present but with a minor role. The T-Stat of the coefficient of the log GDP (table 18.2) indicates that there exists a cointegrating relation between the three I(1) variables at a significant level slightly greater than 5% (the critical value for size 5% is equal to -3.5). This result is conform to that found in the VAR(4) model (figure 18.4) where the acceptation of the hypothesis $r = 1$ is questionable. In this example, the cointegration is less strong than in the example of section 18.3.2. Nevertheless, we keep this model. The equation contains also the growth rates of three hard data (personal consumption, real exports and industrial production index). Moreover, it includes the GDP growth rate lagged with a negative sign expressing that when the growth rate of a given quarter is high then that of the following quarter tends to be reduced. The other regressors (table 18.2) are three survey data and two financial variables (the interest rates spread and the US real stock market index).

Figure 18.4: Rank test in a 5-dimensional VAR(4)

```
Sample:                          1990:01 to 2012:01 (89 observations)
Effective Sample:                1991:01 to 2012:01 (85 observations)
Obs. - No. of variables:         64
System variables:                LPIB LIPI LIMPOR LEXPOR LCONS
Constant/Trend:                  Unrestricted Constant
Lags in VAR:                     4

I(1)-ANALYSIS
 p-r r Eig.Value Trace  Trace* Frac95 P-Value P-Value*
   5  0     0.295 72.748 54.180 69.611   0.027    0.457
   4  1     0.202 43.047 32.500 47.707   0.131    0.588
   3  2     0.135 23.904 18.879 29.804   0.211    0.512
   2  3     0.094 11.621  7.917 15.408   0.178    0.482
   1  4     0.037  3.197  2.760  3.841   0.074    0.097

Trace*, P-Value*: small sample correction (Doornik, 1998)


THE MATRICES BASED ON 1 COINTEGRATING VECTOR:
         LPIB     LIPI    LIMPOR    LEXPOR     LCONS
Beta(1) 1.000  -0.014    -0.138    -0.037    -0.536
         (.NA) (-0.340) (-3.887) (-4.806) (-12.154)


TEST OF EXCLUSION OF LIPI
TEST OF RESTRICTED MODEL:     CHISQR(1) = 0.030 [0.861]
BARTLETT CORRECTION:          CHISQR(1) = 0.017 [0.896]


THE MATRICES BASED ON 1 COINTEGRATING VECTOR:
         LPIB  LIPI   LIMPOR     LEXPOR     LCONS
Beta(1) 1.000 0.000   -0.151    -0.036    -0.521
         (.NA) (.NA) (-10.974) (-4.916) (-18.978)
```

Figure 18.5: Rank test in a 4-dimensional VAR(4)

```
Sample:                          1990:01 to 2012:01 (89 observations)
Effective Sample:                1991:01 to 2012:01 (85 observations)
Obs. - No. of variables:         68
System variables:                LIPI LIMPOR LEXPOR LCONS
Constant/Trend:                  Unrestricted Constant
Lags in VAR:                     4

I(1)-ANALYSIS
 p-r r Eig.Value Trace  Trace* Frac95 P-Value P-Value*
   4  0     0.231 46.966 35.755 47.707   0.059    0.413
   3  1     0.150 24.617 18.806 29.804   0.181    0.518
   2  2     0.086 10.802  8.344 15.408   0.228    0.437
   1  3     0.036  3.140  2.752  3.841   0.076    0.097
```

The adjustment is not very accurate as shown in figure 18.6, especially at the end of the estimation period ($\bar{R}^2 = 0.87$, $SEE = 0.24\%$). The US GDP growth rate is much more volatile than the French one. No unstability has been detected in the US equation as it was in the French case. With real time data, the RMSE lies between 0.3 and 0.35% according to the estimation date.

18.4 Principal components regression

The methodology is described in section 18.4.1. It is illustrated by an example in section 18.4.2 concerning the euro area GDP.

## 18.4.1 Description of principal components regression

This class of models was initiated by Stock and Watson (2002). The modelling involves two steps. First, a large dataset is re-organized into principal components (PC). Second, these PC (or factors) are introduced in a regression, possibly with lags, to explain GDP growth.[8] Moreover, Stock and Watson propose an EM algorithm combined with the PCA to handle the problem of missing observations. Only the statistically significant factors

---

[8]Not all, but the first ones, those which represent a non negligible part of total inertia.

Table 18.2: Estimation of a bridge error correction model for the US GDP

Variables	Transformation	Lag in quarter	Coefficient	T-Stat(*)
GDP (BEA)	Log. level	1	-0.231	-3.3
Personal consumption expenditure (BEA)	Log. level	1	0.121	2.6
Real imports (US census bureau for imports, OFCE for real imports)	Log. level	1	0.040	3.9
GDP (BEA)	Log. diff.	1	-0.312	-4.1
Personal consumption expenditure (BEA)	Log. diff.	0	0.464	6.4
Personal consumption expenditure (BEA)	Log. diff.	1	0.243	2.9
Real exports (US census bureau)	Log. diff.	0	0.078	6.3
Industrial production index (US census bureau)	Log. diff.	0	0.120	3.5
Consumer confidence index (TCB)	Non linear first difference	0	0.736×10^{-5}	3.9
Supplier delivery index (Manufacturers survey, ISM)	First difference	0	0.258×10^{-3}	2.3
NAHB market index	First difference	2	0.221×10^{-3}	3.6
Interest rates spread (Treasury yield 10 years less Treasury bill 3 months)	First difference	1	0.225×10^{-2}	3.4
US real stock index (S&P500)	Log. diff.	3	0.012×10^{-3}	2.5
Constant			0.611	3.7

(*) The T-Stat of I(1) variables follow a non standard distribution. A table is available for the coefficient α in equation (18.2).

Figure 18.6: The US quarterly GDP growth rate observed and fitted (dotted line)
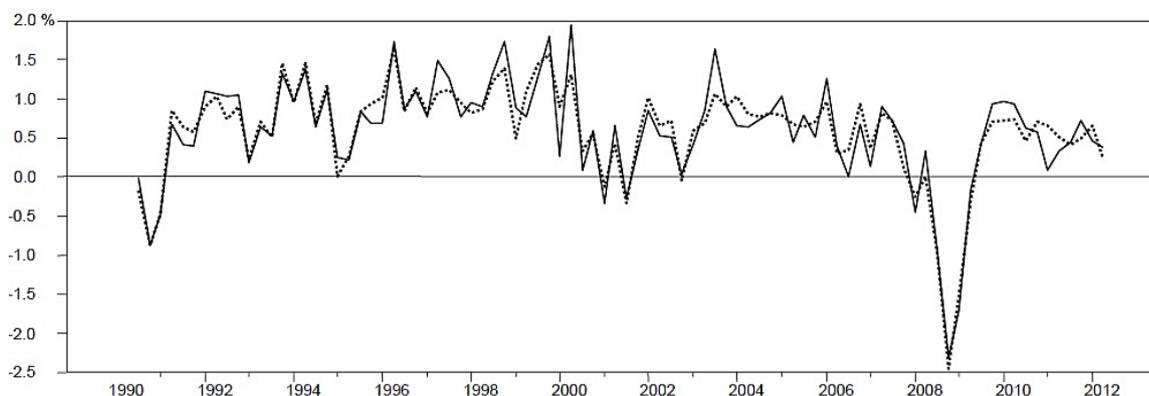
are kept in the regression. The main difference between these models and ours is the size of the dataset. We consider a small dataset instead of the traditional large dataset including all available series, following articles of (Boivin and Ng (2006), Bai and Ng (2008)). Hence we suggest to consider only the targeted series, approximately all series which can help to predict the target but cannot be introduced simultaneously in a regression because of multi-collinearity.

Then, the first step of our approach is the selection of the targeted series to estimate the quarterly GDP growth rate. For that, we choose the algorithm of the Least Angle Regression[9]. This algorithm can be considered as the successor of stepwise algorithms (forward selection, forward stepwise regression). The failure of these latter is that they eliminate too many predictors correlated with the included ones. With the LARS algorithm, all series of the dataset are ranked by decreasing predictive power, according to the selection criterion of this algorithm (see Annex). This criterion does not necessary eliminate series highly correlated, which may well appear in the sorting with adjacent ranks. The ranking given by the LARS algorithm can be exploited either to

[9]Other algorithms are available, see Bai and Ng (2008) for references and comments. On our point of view, the LARS algorithm is the most comprehensive and the most easy to use.

build factors or to select potential regressors in a classical regression.

We propose to use this algorithm but not exactly in the same way that Bai and Ng (2008). A first difference between their implementation and our, is that we introduce series with lags (0, 1, 2, etc.) in our dataset, contrary to the Bai and Ng where only the lag 0 is considered for all series. The reason why lagged variables are present in our dataset is that we expect the LARS algorithm will show the coincident or leading feature of a series. A second difference concerns the inclusion of survey data in our dataset, both in level and in variations,[10,11] We proceed that way because we have frequently observed that either the level of a series or/and its change can help to predict GDP. We expect from the LARS algorithm to give us information about the form to choose for a series (level or change or both). The combination of several lags and several forms multiply the number of variables in the dataset. One advantage of the LARS algorithm is its ability to rank very rapidly a large number of series starting from the most predictive one to the less predictive one according to the selection criterion of the algorithm.

Once the dataset is ranked, it remains to choose the number of targeted series which will be retained to extract principal components. If the objective were to introduce directly these series in a regression, Bai and Ng (2008) have suggested the use of the BIC criterion in order to determine the optimal number of series. But when the objective is to extract factors from these series, they select (without criteria) a fixed number of thirty series, which is well above the optimal number given by the BIC criterion.[12] Note that this optimality corresponds to the in-sample optimality. For an out-of-sample exercise, it is better to base estimation on a greater number of series in order to make estimation less dependent on extreme changes in individual series.

The choice of the number of targeted series to take into account remains an issue insofar as only an empirical investigation is able to guide this choice. The BIC information criterion gives an indication of the minimum number of series to be included but our empirical experience leads us to increase slightly this number. When this minimum number is found to be very low, implementing a bridge model with these few selected series seems a good option as compared to a factor model.

Once the small dataset is determined, the principal components (PC) extraction is carried out. It seems preferable that this extraction be based only on released data. Consequently, when some hard data have to be forecast to cover the coincident quarter, these forecasts cannot be taken into consideration in the factor determination. They will only serve to extrapolate factors in order to compute the GDP estimate. Then, the PC are introduced in the GDP equation and the significant ones are kept. It is not necessary to lag the PC because each series composing a PC has been introduced with its appropriate lag.

The PC extraction as well as the principal component regression has to be reconsidered at each estimation date. Note that a factor which bears the number k for an estimation date, can be the $(k-1)$ or the $(k+1)$ factor for another date. This may occur for factors representing a low part of total inertia.

18.4.2 Estimation of the euro area GDP growth by a factor model

The initial dataset contains survey data (Industry, Consumers, Construction, Retail trade), hard data in volume (industrial production excluding construction, construction production, exports and imports, retail sales and unemployment rate) and other data (the real dollar/euro exchange rate, the real effective exchange rate, the real oil price, the real stock market index, the short and long term interest rates and the interest rate spread. These data are transformed if necessary to reach stationarity. Each series can be included several times according to its lag (0,1,...) and to its transformation (level, change and non linear change).

Then the LARS algorithm is run to rank the initial dataset accordingly to the target, here the quarterly GDP growth rate. After an empirical investigation, we retain the first fifteen series of the ranking, listed in table

---

[10]Linear Δz and non linear $\Delta z \times |\Delta z|$.

[11]In this case, the level must be stationary.

[12]Their initial dataset contains about 120 series.

Table 18.1: The first fifteen series of the ranking given by the LARS algorithm (end 2011)

Variables by decreasing order	Transformation	Lag in quarter
Industrial production excluding construction (Eurostat)	Growth rate	0
Unemployment rate (Eurostat)	Log. diff.	0
Real exports (Eurostat for exports, OFCE for real exports)	Growth rate	0
Construction production (Eurostat)	Growth rate	0
Retail sales (Eurostat)	Growth rate	0
Major purchases intentions, next 12 months (Consumer survey, DGECFIN)	Non linear first difference	1
Major purchases intentions, next 12 months (Consumer survey, DGECFIN)	First difference	1
Opinion on the present business situation (Retail trade survey, DGECFIN)	First difference	0
Employment expectations (Retail trade survey, DGECFIN)	Non linear first difference	0
Opinion on activity (Construction survey, DGECFIN)	First difference	3
Major purchases intentions, next 12 months (Consumer survey, DGECFIN)	Level	1
Export order-book (Industrial survey, DGECFIN)	Non linear first difference	0
Opinion on stocks (Retail trade survey, DGECFIN)	Level	1
Opinion on stocks (Retail trade survey, DGECFIN)	Level	2
Real Dollar/euro exchange rate	growth rate	2

18.1.[13]

The first five are hard data, followed by nine soft data (some are leading, other coincident) and, at the fifteenth place, the real dollar/euro exchange rate. Note that the ranking can differ slightly according to the dates at which it is carried out. But generally, among the first twenty series, we find the same series although not always in the same order. We usually check the ranking once a year.[14]

The fifteen principal components are extracted from this small dataset and introduced as regressors in the GDP equation. Seven factors are found significant over the period 1992Q2–2012Q1 (number: 1, 2, 3, 4, 7, 8, 11) and are retained to estimate the quarterly GDP growth rate. If we look at the most recent in-sample adjustment, plotted on Figure 3, it is very accurate ($\bar{R}^2 = 0.93$, SEE=0.16%). With real time data, the RMSE is greater than the SEE and lies between 0.2 and 0.25% according to the estimation date.

18.5 Forecasting missing data

We describe in Section 18.5.1 our methodology and give an example in Section 18.5.2

---

[13]We successively retain 20, 19, 18, . . . 10 series and select the best choice to estimate GDP.
[14]In this example we use a ranking obtained at the end of year 2011.

Figure 18.7: The euro area quarterly GDP growth rate observed and fitted (dotted line)
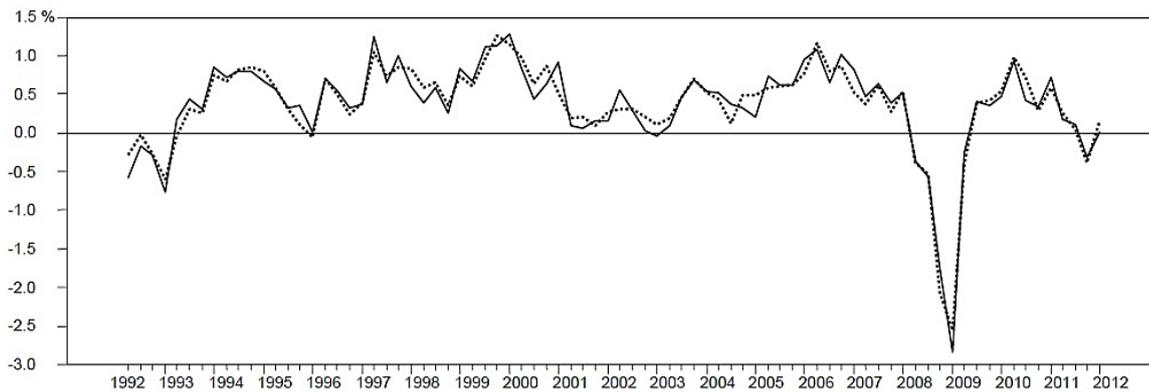
18.5.1 Methodology

In general it is a little after the release of GDP for the quarter $(T-1)$ that one seeks to estimate the quarter T and possibly the quarter $(T+1)$.[15] At this date, no monthly hard data is available on the coincident quarter T, and the last known data is the last month of the quarter $(T-1)$. Concerning survey and financial data, two months on quarter T are yet available. For survey data, one can without appreciable loss of accuracy assimilate the quarter to the average of the two available months. As for the financial variables, they are delayed when they enter models and therefore it is useless to predict them. Only hard data have to be forecast for the three months of quarter T using their own past, survey variables present and past, and past financial variables. One month later, hard data have to be forecast for two months ahead. For survey and financial data, the three months of the quarter are known. And finally about two weeks after, two months of hard data are known. Thus, the forecasting work concerns only the hard data and, at most, for a 3-month horizon.

If we take again the examples mentioned in sections 18.3.2 and 18.3.3, one have to forecast the four series (industrial production index, real consumption, real exports, real imports) that are not linked by a co-integrating relation.[16] To do this, we can estimate a system of four equations, these variables being clearly interdependent. We think first to a VAR model. In fact this model is not adapted to our case, because if interdependencies exist between the four variables they are instantaneous, the links not instantaneous to each equation.[17,18] These are not necessarily the same survey (or financial) variables that will be found significant in each equation. Finally, these equations will have little regressors in common. For all these reasons, the appropriate estimation method is one that applies to the Seemingly Unrelated Regressions (SUR) model.

Rather than using specific variables in each equation as described above, we may employ factors, constructed according to the method of section 18.4. For each endogenous, targeted predictors are selected according to the LARS algorithm. A small number of them are retained and will form the factors. Finally the system of equations is estimated taking into account the instantaneous correlations between the equation errors and keeping only the significant factors of each equation.

18.5.2 Forecasting the euro area IPI monthly growth by a factor model

The initial dataset contains survey data (Industry, Consumers, Construction, Retail trade), and other data (the real dollar/euro exchange rate, the real effective exchange rate, the real oil price, the real stock market

---

[15] Few days after for the USA and two weeks after for the euro area.

[16] If they were, correction terms could be introduced into the systems we propose.

[17] Between one variable in t and the other in $t-h$, $h \geq 1$. being infrequent. Moreover, we plan to add survey and financial variables.

[18] Released one or two months before.

index, the short and long term interest rates and the interest rate spread). These data are transformed if necessary to reach stationarity. Each series can be included several times according to its lag (0,1,. . .) and to its transformation (level, change and non linear change).

Then the LARS algorithm is run to rank the initial set according to the target, the monthly IPI growth rate. We choose the first twelve series of the ranking, listed in table 18.1. Eight of twelve are in non linear first difference, a transformation appropriate to high volatility. Seven belongs to the industrial survey.

The twelve principal components are extracted from the small dataset listed in table 18.1. The IPI equation, which gives the IPI monthly growth rate, contains three autoregressive terms and two factors (numbers 1 and 3). The most recent in-sample adjustment is not accurate ($\bar{R}^2 = 0.55$, SEE=0.70%) as the series is very volatile. Note that if we used an AR(3) model, we would have found $\bar{R}^2 = 0.16$, SEE=0.96%.

Table 18.1: The first twelve series of the ranking given by the LARS algorithm (end 2011)

Variables by decreasing order	Transformation	Lag in month
Confidence indicator (Industrial survey, DGECFIN)	Non linear first difference	1
Order-book (Industrial survey, DGECFIN)	Non linear first difference	1
Export order-book (Industrial survey, DGECFIN)	Non linear first difference	0
Sentiment indicator (DGECFIN)	Non linear first difference	1
Confidence indicator (Industrial survey, DGECFIN)	Non linear first difference	2
Production expectations (Industrial survey, DGECFIN)	First difference	1
Unemployment next 12 months (Consumer survey, DGECFIN)	Non linear first difference	1
Unemployment next 12 months (Consumer survey, DGECFIN)	First difference	0
Employment expectations (Industrial survey, DGECFIN)	Non linear first difference	0
Order-book position (Construction survey, DGECFIN)	First difference	2
Confidence indicator (Industrial survey, DGECFIN)	Non linear first difference	0
Major purchases intentions, next 12 months (Consumer survey, DGECFIN)	Non linear first difference	1

18.6 Mixed-data sampling (MIDAS) models

The objective of these models is to work with both frequencies in the GDP equation and to use only the monthly released information without predicting the missing monthly data on the quarter to forecast. If t indicates a particular quarter and m a month within that quarter ($m = 1, 2, 3$), let X be the level of a monthly variable and X_t its quarterly value. We have for example for a conversion by averaging:

$$X_t = \frac{1}{3} \sum_{m=1}^{3} X_{t,m}$$

(18.1)

Then monthly and quarterly first differences are linked by (18.5)

$$\Delta X_t = \frac{1}{3}\Delta^m X_{t,3} + \frac{2}{3}\Delta^m X_{t,2} + \Delta^m X_{t,1} + \frac{2}{3}\Delta^m X_{t-1,3} + \frac{1}{3}\Delta^m X_{t-1,2} \qquad (18.2)$$

where Δ^m is the monthly first difference operator, i.e. $\Delta^m X_{t,3} = X_{t,3} - X_{t,2}$ $\Delta^m X_{t,2} = X_{t,2} - X_{t,1}$ $\Delta^m X_{t,1} = X_{t,1} - X_{t-1,3}$

Concerning the levels in log, we have approximately[19]:

$$Log(X_t) \approx \frac{1}{3}\sum m = 1^{m=3} Log(X_{t,m}) + Log 3 \qquad (18.3)$$

Let $x_{t,m} = Log(X_{t,m})$ and $x_t = Log(X_t)$. By first differencing (6), we deduce from (5) the link between monthly and quarterly growth rates:

$$\Delta x_t = \frac{1}{3}\Delta^m x_{t,3} + \frac{2}{3}\Delta^m x_{t,2} + \Delta^m x_{t,1} + \frac{2}{3}\Delta^m x_{t-1,3} + \frac{1}{3}\Delta^m x_{t-1,2} \qquad (18.4)$$

In all cases (18.4), (18.5) and (18.7), the quarterly level/first difference/growth rate are the weighted average of the monthly level/first difference/growth rate. Equations (18.4) and (18.5) serve for survey data whereas (18.7) for hard data.

In a MIDAS equation, the regressors will be the variables of the right-hand side of equations (18.4), (18.5) or (18.7) according to the monthly indicator (level/first difference/growth rate). For that, each month constitutes a quarterly series. For example, in case of growth rate, the second month leads to the quarterly series $\{\dots, \Delta^m x_{t-1,2}, \Delta^m x_{t,2}, \Delta^m x_{t+1,2}, \dots\}$ Thus to each monthly indicator is associated three quarterly series.

Now in order to explain what is a MIDAS model, take a simple example, that of a bridge equation with a single explanatory variable. Let Δy denote the GDP growth and (18.8) the bridge model in quarterly frequency

$$\Delta y_t = \alpha \Delta x_t + \beta \qquad (18.5)$$

The corresponding MIDAS model is written according to (7)

$$\Delta y_t = a_0 \Delta^m x_{t,3} + a_1 \Delta^m x_{t-1,3} + b_0 \Delta^m x_{t,2} + b_1 \Delta^m x_{t-1,2} + c_0 \Delta^m x_{t,1} + d \qquad (18.6)$$

In fact, we have said that only the monthly released information was used. So, when two months of the coincident quarter are known, equation (10a) is estimated and, when one month of the coincident quarter is known, equation (10b) is considered .

$$\Delta y_t = a_1 \Delta^m x_{t-1,3} + b_0 \Delta^m x_{t,2} + b_1 \Delta^m x_{t-1,2} + c_0 \Delta^m x_{t,1} + d \qquad (18.7a)$$
$$\Delta y_t = a_1 \Delta^m x_{t-1,3} + b_1 \Delta^m x_{t-1,2} + c_0 \Delta^m x_{t,1} + d, \qquad (18.7b)$$

Equations (10a) and (10b) correspond to unrestricted MIDAS models, where the coefficients are estimated unrestrictedly. But collinearity problems can emerge between the different months of a given variable, especially when survey series are used in level. Thus, the form of the coefficient distribution must be parameterized. For example, suppose that the GDP growth is explained by a survey indicator in level and that the bridge model in quarterly frequency is

$$\Delta y_t = \alpha_0 \Delta X_t + \alpha_1 \Delta X_{t-1} + \beta \qquad (18.8)$$

---

[19]First order.

When two months of the survey indicator are known on the coincident quarter, the corresponding MIDAS model is written

$$\Delta y_t = b_1 X_{t,2} + b_2 X_{t,1} + b_3 X_{t-1,3} + b_4 X_{t-1,2} + b_5 X_{t-1,1} + c \qquad (18.9)$$

The coefficients $b_h(h = 1, ..., 5)$ cannot be estimated unrestrictedly and must be parameterized as in distributed lag models. The usual lag distribution is the exponential Almon lag, sufficiently flexible to describe many kinds of lag distributions, defined with few parameters entering in a non linear manner in the weights. For example, when the parameterization depends on two coefficients $a1$ and $a2$, the lagged coefficients b_h are given by equation (18.13):

$$\frac{exp(a_1 h + a_2 h^2)}{\sum h = 1^H exp(a_1 h + a_2 h^2)} \qquad (18.10)$$

Non linear least squares are then used to estimate coefficients a_1 and a_2 in the GDP regression.

On our point of view, one cannot expect a significant improvement in results using a MIDAS model. We know that a weighted mean is not very sensitive to its weights. So if you replace identical weights by others as in (18.12), the results are little changed. However it is a way to solve the problem of missing monthly data on a coincident quarter. Is it superior to the use of auxiliary equations? It is difficult to answer this question as MIDAS models proposed in the literature are very simple. Most often, only one regressor enters the GDP equation, see for example Clements and Galvao (2008). Even in this case, the results presented in this paper (table 18.2) are not convincing.

18.7 Conclusion

All models described in this chapter are easy to implement and lead to forecasts the accuracy of which is as good as or better than forecasts carried out with more sophisticated models. However, until now, the accuracy of sophisticated models is difficult to appreciate because very few are really in use relatively to the huge number of papers published on the subject. Sophisticated models such that presented in Chapter 16 are elegant because they handle the problem of missing observations without resorting to auxiliary equations and the problem of mixed frequency data. On our point of view their drawbacks are (i) the use of non targeted datasets either to forecast the GDP or the missing observations of monthly indicators (ii) the poor dynamic, only generated by factors, which ignores the character coincident or leading of each monthly series (iii) the impossibility to find the source of forecast errors. The advantage of our simple models is that when you make a big forecast error, you can explain why, which may allow you to improve the model. The drawback is the lack of elegance and the old fashioned aspect.

Annex

18.A LARS algorithm: A short description

This is the algorithm of the Least Angle Regression, Efron et al. (2004). We do not use it to run Least Angle Regression but only to rank a dataset given a target.

This algorithm can be considered as the successor of stepwise algorithms (forward selection, forward stepwise regression). In the forward selection regression, the $(k+1)th$ series entering the selection is the one exhibiting the highest correlation (in absolute value) with the residual of the regression on the k series of the previous selection and the magnitude of the step is given by this correlation. This selection method tends to be too aggressive in the sense that it eliminates too many predictors correlated with the included ones. To overcome this drawback, the forward stepwise regression reduces the magnitude of the step in the previous progression. But this new procedure appears to be a particular case of the LARS algorithm without presenting any specific advantages. In the LARS algorithm, the $(k+1)th$ series entering the selection, together with the magnitude of the step, are determined endogenously such as the spread between the series and its estimate at the step $(k+1)$ has the same correlation with the series already included in the selection. Geometrically a correlation is an angle and the progression (equi-angle) is realized according to a smaller angle (least angle) than this chosen by the classical forward selection. The ranking given by the LARS algorithm can be used either to build factors or to select potential regressors in a classical regression. We give below a brief description of the algorithm.

18.A.1 Definition of an equiangular vector

One assumes that p vectors (variables) are linearly independent and standardized. Let Z be the matrix formed with the p vectors $Z = (x_1...x_p)$ and let e be a p-vector of one. The equiangular vector u is the unit vector making equal angles, less than 90%, with the columns of Z. It is defined by equation (A1):

$$Z'u = ke || u = 1 u = kZ(Z'^{-1}ek = (e'Z(Z'^{-1}e)^{-1/2} \tag{A1}$$

If $p = 2$, u is the bisector vector of (x_1, x_2)

18.A.2 Notations

Let y be the target and let X be the matrix formed with the n standardized variables $X = (x_1 \ldots x_n)$.

At step j:

- let $\hat{\mu}_j$ be the current LARS estimate obtained with j variables,
- let $c(\hat{\mu}_j) = x'_j(y - \hat{\mu}_j)$ a scalar proportional to the correlation between estimation residual and variable xj ,
- let $c(\hat{\mu}j) = X'(y - \hat{\mu}_j)$ be the associated vector.

To simplify, one assumes hereafter that correlations are positive.

18.A.3 Iterations

The LARS algorithm is an iterative technique that begins with $\hat{\mu}_0 = 0$. The correlation vector $c(\hat{\mu}_0) = X'(y - \hat{\mu}_0)$ is calculated and the first entering variable is that which has the highest correlation with y, say x_1, thus $max_j c_j(\hat{\mu}_0) = c_1(\hat{\mu}_0)$.

STEP 1

Step 1 describes how the new LARS estimate $\hat{\mu}_1$ is formed. LARS augments $\hat{\mu}_0$ in directions of x_1 (equation A2)

$$\mu'_1(\gamma) = \hat{\mu}_0 + \gamma u_1(u_1 x'_1) \tag{A2}$$

Classic forward regression would choose γ large enough to make μ_1 equal to the projection of y into the space generated by x_1. LARS uses an intermediate value of γ, the value that makes the residual $(y - \mu_1)$ equally correlated with x_1 and the new entering variable. In equation (A3) we see that the current correlations depend on γ :

$$c_j[\mu_1(\gamma)] = x'[y - \mu_1(\gamma)] \tag{A3}$$

and we note that $c_1[\mu_1(\gamma)] \leq c_1(\hat{\mu}_0$ because $c_1[\mu_1(\gamma)]$ is a decreasing function of γ.

The algorithm determines together the smallest γ (say γ_1) and the entering variable by solving equation (A4):

$$c_1[\mu_1(\gamma)] = c_j[\mu_1(\gamma)]j \neq 1 \tag{A4}$$

Suppose that x_2 is the entering variable and let $\hat{\mu}_1$ be $\hat{\mu}_1 = \mu_1(\gamma_1)$, then $c_1(\hat{\mu}_1) = c_2\hat{\mu}_1$.

Finally, one computes the equiangular vector u_2 for (x_1, x_2) with equation (A1).

STEP 2

Step 2 describes how the new LARS estimate $\hat{\mu}_2$ is formed. LARS augments $\hat{\mu}_1$ in directions of x_2 (equation A5)

$$\mu_2(\gamma) = \hat{\mu}_1 + \gamma u_2 \tag{A5}$$

One determines together the smallest γ (say γ_2) and the entering variable by solving (A6)

$$c_1[\mu_2(\gamma)] = c_2[\mu_2(\gamma)] = c_j[\mu_2(\gamma)] \tag{A6}$$

Suppose that is x_3 the entering variable and let $\hat{\mu}_2$ be $\hat{\mu}_2 = \mu_2(\gamma'_2)$, then

$$c_1(\mu_2) = c_2(\mu_2) = c_3(\mu_2).$$

Finally, one computes the equiangular vector u_3 for (x_1, x_2, x_3) with equation (A1).

Etc

If $n = 3$ then one takes $\hat{\mu}_3 = \hat{y}_3$ the projection of y on $L(x_1, x_2, x_3), \mu L x$.

Bibliography

Bai J., Ng S. (2008), Forecasting Economic Time Series Using Targeted Predictors, Journal of Econometrics 146, 304-317.

Boivin J., Ng S. (2006), Are more data always better for factor analysis?, Journal of Econometrics, vol 132, issue 1, May, 169-194.

Clements M.P., Galvao A.B. (2008), Macroeconomic forecasting with mixed-frequency data: Forecasting US output growth, Journal of Business and Economic Statistics, 26, 546-554.

Efron B., Hastie T., Johnstone I., Tibshirani R. (2004), Least Angle Regression, Annals of Statistics 32:2, 407-499.

Ericsson N.R., MacKinnon J.G. (2002), Distributions of error correction tests for cointegration, Econometrics Journal, vol. 5, pp. 285-318.

Ghysels E., Santa-Clara P., Valkanov R. (2005), There is a risk-return trade-off after all, Journal of Financial Economics, 76(3), 509-548.

Ghysels E., Santa-Clara P., Valkanov R. (2006), Predicting volatility: Getting the most out of return data sampled at different frequencies, Journal of Econometrics, 131(1-2), 59-95.

Johansen S. (1995), Likelihood-based inference in cointegrated vector auto-regressive models, Oxford University Press. Stock,

Stock J.H., Watson M.W. (2002), Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association, 97, 1167-1179.

VI Evaluation and Dissemination of Cyclical Composite Indicators

19

How did the OECD Composite Leading Indicators Perform?

Authors

Roberto Astolfi
OECD
roberto.astolfi@oecd.org

Emmanuelle Guidetti
OECD
emmanuelle.guidetti@oecd.org

Michela Gamba
OECD
michela.gamba@oecd.org

Pierre-Alain Pionnier
OECD
pierre-alain.pionnier@oecd.org

Table of Contents

19.1 Introduction

This chapter discusses the performance of the OECD System of Composite Leading Indicators (CLIs) in the run-up to, and during, the Great Recession.[1] The OECD CLIs, first developed in the 1970s, are qualitative indicators designed to early detect turning points in economic activity. They anticipate turning points by combining into a robust and timely signal the information that can be derived from a set of leading components (OECD (2012)).

The performance of the OECD CLIs is assessed here using real-time analysis. Real-time analysis is based on the provisional and partially revised data that were actually available at the time of the compilation and dissemination of the relevant CLIs. The analysis evaluates the ability of OECD CLIs to announce the peak and the subsequent trough of the Great Recession in the G7 countries, and the extent to which the initial signal has been maintained. Results suggest that the CLIs were able to anticipate the Great Recession in the G7 countries some months in advance, although, by their very nature, they could not give an indication on the depth of the coming crisis. Our results confirm the conclusions previously reached by Gyomai and Guidetti (2011) who relied on an *ex-post* analysis.

The remainder of this chapter is organised as follows: Section 19.2 provides a synthetic overview of the timing of the detection of turning points and their subsequent announcement to the public; Section 19.3 presents the results of the real-time analysis of the CLIs' performance during the Great Recession; section 19.4 concludes. An annex is provided at the end of this chapter.

19.2 When were the turning points of the Great Recession identified and first announced?

In September 2007, the CLI for the OECD area as a whole had recorded a significant decline, which suggested that a possible deterioration in economic activity was approaching.[2] Hence, the OECD CLI Press Release headlined "*moderating outlook*". Over the following months, the CLIs confirmed the initial symptoms. "*Weakening outlook*" and "*Continued weakening outlook*" were the headings of the Press Releases in the last quarter of the year. In January 2008, the signal further worsened and the message was therefore turned into "*downswing*". In the subsequent months, the reading of CLI evolutions sharply declined to eventually reach levels as low as those seen during the Oil Crisis in the 70s ("*Lowest level since 70s*" in February 2009) and even lower than that, with "*New low*" announced in March 2009. Figure 19.1 illustrates the evolution of the headings.[3] On the recovery side, the CLI identified the first signs of a likely improvement in economic activity in May 2009. On that occasion, the OECD headlined "*strong slowdown in the OECD area but the pace of the deterioration is easing.*"

In both cases, at peak and trough, the OECD was able to signal the approaching turning points thanks to

---

[1]This chapter draws extensively upon the authors' working paper "The use of short-term indicators and survey data for predicting turning points in economic activity: A performance analysis of the OECD system of CLIs during the Great Recession" (Astolfi et al. , 2016).

[2]Until April 2012, the Index of Industrial Production (IIP) was used as the reference for operational purposes. This reflected the fact that real GDP figures needed to quantify the reference business cycle were available on a quarterly basis for only half of the member countries (and initially none was available on a monthly basis). Instead, the IIP was available for all OECD member countries on a monthly and quarterly basis (Fulop and Gyomai (2012)). Of note is also the fact that the IIP represents the most "cyclical" component of GDP, accounting for approximately 35% share of gross value added in the mid-1980s. The industrial sector, being a significant consumer of services activities, also drives supply in a significant part of the private service sector. Since April 2012, in response to improvements in national statistical information systems (all OECD countries now produce quarterly estimates of GDP) and because of the industrial sector's diminishing share of total GDP (direct and indirect) in recent decades in most OECD economies, the CLI system switched to using GDP as the reference series.

[3]See Annex in section 19.A for further details on how quantitative scores are assigned to the headings of CLI Press Releases.

Figure 19.1: Evolution of CLIs Press Releases headlines during the Great Recession, OECD area

Note: The vertical lines identify the turning points detected by the CLIs for the OECD area as a whole (peak in June 2007 and trough in February 2009, marked in red) and GDP (marked in dotted black, with a peak in December 2007 and a trough in May 2009).
Source: CLI Press Release, OECD.

the continuous monitoring of CLI growth rates, which initially recorded a significant reduction and then turned negative. With hindsight, the CLI for the OECD area as a whole peaked in June 2007, seven months ahead of the corresponding peak for GDP, which took place in December 2007.[4] Similarly, ex-post data suggest that the CLI for the OECD area troughed in February 2009 while GDP reached its minimum in May 2009 (both CLI and GDP turning points are included in Figure 19.1).

The remainder of this paper addresses two questions based on the experience of the Great Recession:

(i) Does the OECD System of CLIs sufficiently allow for the timely detection of turning points? and

(ii) How stable is the location of turning points in economic activity?

To address these two questions, we use both ex-post and real-time analysis. While the former is based on the information available today, the latter only uses the historically available data.

19.3 Turning point detection

Before reporting on the results of the ex-post and real-time analysis, we briefly present a chronology of events related to the detection of the turning points of both the CLIs and the reference series. This allows clarification of what exactly is measured when assessing the performance of the CLIs in the ex-post and real-time analysis.

---

[4]Based on CLIs as available in September 2014.

19.3.1 Chronology of events

The exact dating of turning points requires some time after they manifest. This is because any dating algorithm necessitates a certain number of observations after the turning point to be able to single out a maximum (peak) or a minimum (trough) in the time series (distance a in figure 19.2).[5] Nonetheless, a constant monitoring of the evolution of CLI growth rates allows the OECD to anticipate the formal identification of the turning points and therefore to announce to the public the possibility that a turning point is approaching (distance b in figure 19.2).

Figure 19.2: Chronology of events
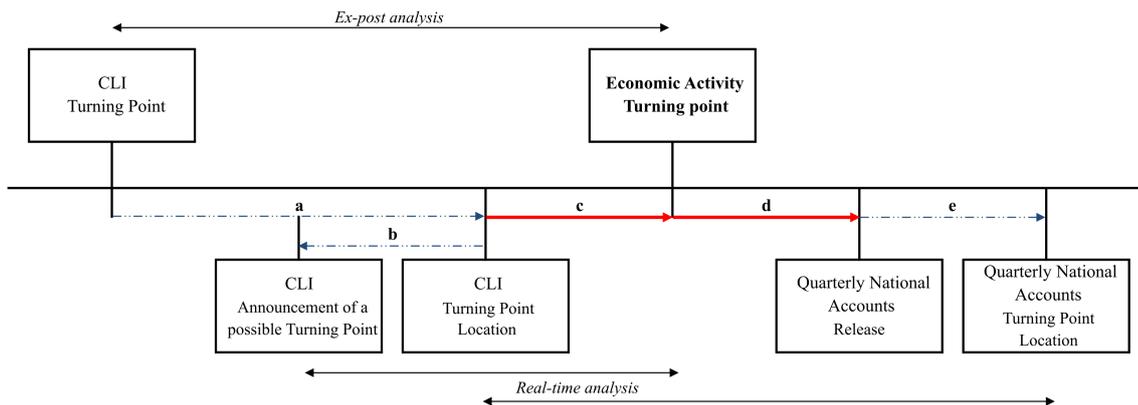
Source: Our elaboration.

The early announcement of approaching peaks or troughs may anticipate by a certain number of months the beginning of recessions and recoveries in economic activity (distance $b + c$ in figure 19.2). However, not only recessions and recoveries in economic activity can be observed after Quarterly National Accounts (QNA) are released (distance $b + c + d$ but, as in the case of the CLIs, their formal identification requires time after they manifest (distance e).

Hence, the *ex-post* analysis evaluates to what extent CLI turning points, located with hindsight using all the information available today, lead peaks and troughs in the economic activity (distance $a + c$). On the other hand, real time analysis is interested in (1) the lag between the initial announcement of a possible turning point and the date at which it occurs in economic activity (distance $b + c$); and (2) the interval between the formal identification of turning points in the CLI and GDP (distance $c + d + e$).

19.3.2 Assessment of CLI leading properties based on the latest CLI vintage

The *ex-post* analysis shows that CLIs anticipated turning points in GDP for all G7 countries both at the onset and at the end of the Great Recession. At the beginning of the crisis, the CLI for the United States, for instance, peaked in June 2007, 4 months before GDP had reached its own peak. In March 2009, the US CLI anticipated a trough that also correctly led to the following trough in US GDP. In the latter case, however, the leading period was shorter as the CLI and the GDP troughs were only two months apart. For more details of this ex-post analysis, reference is made to Gyomai and Guidetti (2011)).

---

[5]The OECD employs a simplified version of the Bry and Boschan (1971) Bry-Boschan algorithm to date the turning points (Gyomai et al. (2017)).

19.3.3 Real-time analysis

Results presented in the previous section on the performance of the G7 CLIs during the Great Recession need to be complemented with a real-time analysis. Indeed, the ex-post analysis is based on the current set of CLIs which may give a too favourable picture of their historical performance in signalling fluctuations in the reference series. This is because currently available CLIs are evaluated after the underlying data have been revised and more information has become available. Diebold and Rudebusch (1991) were among the first to make this claim and to show that the forecasting performance of the Index of Leading Economic Indicators (LEIs) released by the Conference Board in the US deteriorated significantly in a real-time framework.

From their initial estimate to their latest release, the OECD CLIs may undergo both regular and exceptional revisions:

- Regular revisions of the CLIs can be the result of a revision of the components by National Statistical Offices[6] or can be due to the filtering process once new data points become available. Since filters (extraction of the cycle, seasonal and trading-day adjustment, and outlier detection) operate on the whole sample, the inclusion of a new data point may produce a revision of the entire time series.

- Beyond the regular revision process, exceptional revisions can occur due mostly to the implementation of new methodologies. For example, in December 2008 the OECD replaced the Phase Average Trend (PAT) method with the double Hodrick Prescott (HP) filter in order to extract the cyclical component of time series (Nilsson and Gyomai , 2011). Revisions can also be ascribed to changes in the set of the indicators which are used to compile a CLI. Indicators may be replaced should their performance deteriorate over time and new series can be added to capture structural changes in the economy.

The occurrence of regular revisions and changes in the methodology of CLIs are strong arguments in favour of a real-time performance analysis. The OECD publishes on its website all CLI vintages since the end of the 1990s or the beginning of the 2000s[7], depending on the country. This limits the real-time performance analysis to the last 15 years, but the Great Recession is fortunately covered in the real-time dataset.

In what follows, real-time data are used to assess the ability of the OECD CLIs to identify the peak and the trough during the Great Recession for all G7 countries by: (1) verifying that the location of those turning points has remained stable from one CLI monthly release to the next; and (2) determining at what point in time peaks and troughs have been actually identified.

Stability of turning points in real time

A real-time assessment of the OECD CLIs' performance during the Great Recession shows that the location of CLI turning points remains broadly stable over time. From one release to the next, turning points generally remain in a 3-month corridor, a result that can be considered fully satisfactory for economic policy purposes. For example, figure 19.3 below shows that the June 2007 peak and the March 2009 trough identified for the United States remain very stable from one vintage to the next. In January 2008, it had been estimated that the CLI had reached a peak in June 2007. After the initial estimate, the date of the peak has barely been revised (see blue line). A rather stable pattern can also be observed for the trough anticipating the possible end of the crisis. Initially located in September 2009, the date at which the CLI troughed has shifted a month ahead in mid-2011 and has since remained unchanged (see red line).

---

[6]Although, as mentioned above, the impact of these revisions is limited since an important criterion for the selection of indicators entering the construction of the CLIs is precisely that they should not be subject to significant revisions.

[7]No electronically-supported files are available for earlier CLI releases.

Figure 19.3: Stability over vintages of turning points in the US CLI at the time of the Great Recession
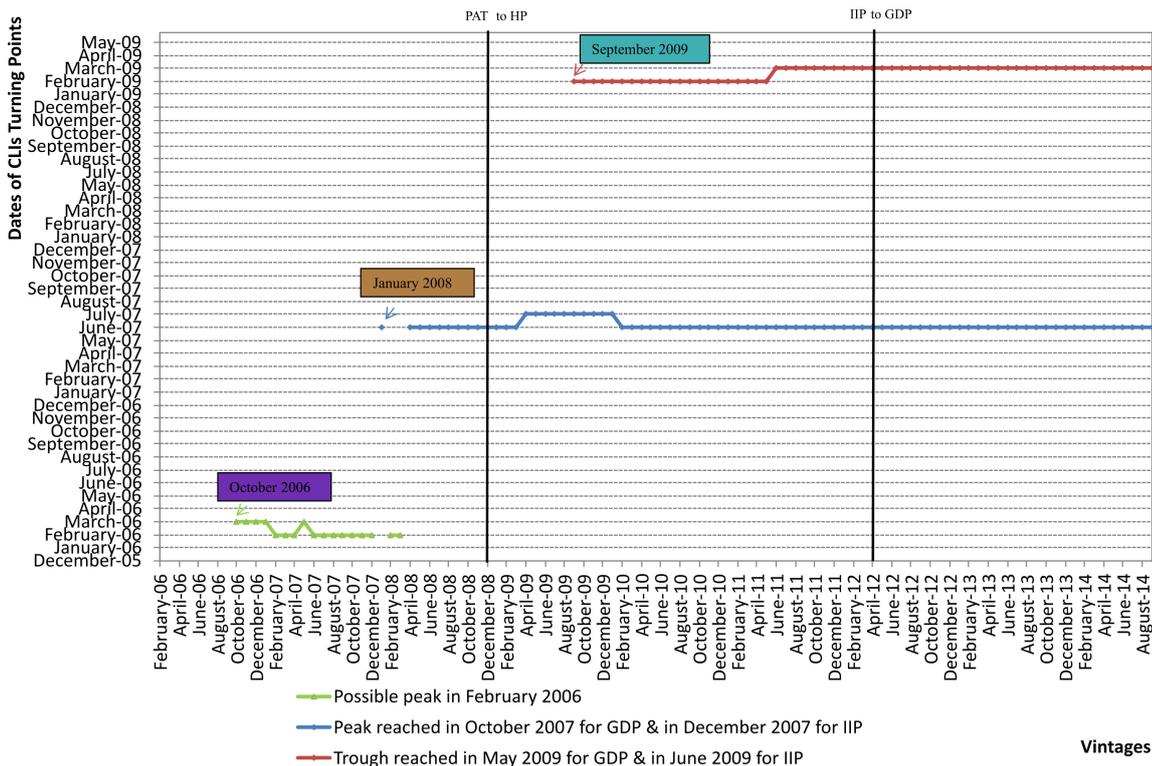
Note: The horizontal axis refers to CLI vintages and the vertical axis to the dates of CLI turning points. For instance, the blue line shows that the June 2007 peak in the current US CLI (see vertical axis) was first detected in January 2008 (see horizontal axis) and was already located in June 2007 at that time. US GDP reached a peak in October 2007, as indicated in the legend.
Source: Main Economic Indicators, OECD.

The revision analysis of the sign of CLI growth rates over the last three months provides additional insights on the stability of turning-point dating. In practice, this is what the OECD does to interpret the latest CLI results for the monthly Press Releases. Figure 19.4 shows the stability of signs of the CLI evolution over the last three months.

Figure 19.4: Signs of the US CLI evolution over 3 months (by date and vintage)
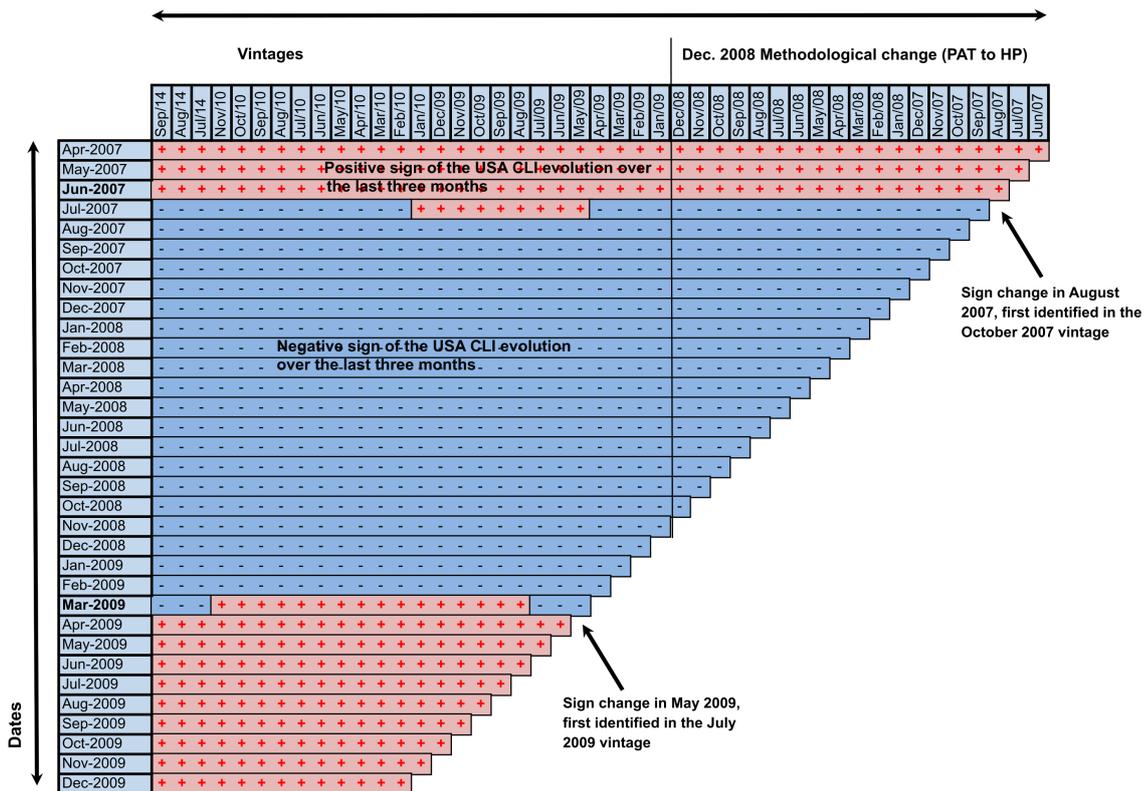
Note: _The horizontal axis refers to CLI vintages and the vertical axis to time. Periods where the 3-month CLI evolution is negative are marked in blue, those where the 3-month CLI evolution is positive are marked in red. What can be derived from the figure is that the US CLI evolution over 3 months turned negative in August 2007 (see vertical axis). This sign change was first detected in the October 2007 release of the US CLI (see horizontal axis). The 3-month evolution of the latest US CLI considered here (September 2014 vintage: first column) also changes sign in August 2007 (see vertical axis). Source: Main Economic Indicators, OECD._

Of particular note is the fact that in October 2007 (see the relevant "vintage" in figure 19.4), the 3-month evolution of the US CLI showed a change for the first time, indicating that the CLI had started declining back in August 2007 (see the relevant "date" in figure 19.4). The message was then confirmed in November 2007 and, on that occasion, the OECD announced a "*Possible downturn in the United States*", two months in advance compared to the formal identification of the peak (January 2008).

Timeliness of the formal identification of turning points

The formal identification of turning points requires time after their occurrence because dating algorithms necessitate a certain number of observations after the turning point to be able to single out a maximum (peak) or a minimum (trough) in the time series. This is not specific to the Bry and Boschan (1971) dating algorithm employed by the OECD, as indeed all turning point detection techniques entail some time to extract the underlying signal from noisy data. Table 19.1 and table 19.2 below report the dates of occurrence of turning points for CLIs (column A) and GDPs (column B) as formally located by the OECD dating algorithm, as well as the date when CLI turning points were first formally identified in real time (column C) and the GDP release date by national accounts (column D).

Before interpreting table 19.1 and table 19.2, it is important to remember that the real-time performance of the

OECD CLIs can be assessed in two different ways, as indicated in the chronology of events (figure 19.2):

(i) dates when CLI turning points are identified can be compared to dates when corresponding turning points in GDP actually occur, or

(ii) they can be compared to dates when turning points in GDP can be identified based on national accounts' data.

With only a few exceptions, CLI turning points during the Great Recession could be formally identified approximately six months after they had occurred (difference between columns A and C in table 19.1 and table 19.2). Note that this additional delay implies that CLI turning points could not always be formally identified before turning points in GDP occurred. In other words, distance c in the chronology of events could sometimes be reduced to zero or become slightly negative. This was especially the case for the 2009 trough. Nevertheless, two additional points are worth noting here:

- Thanks to informal identification based on the sign of CLI growth rates, the OECD was able to announce CLI turning points before they could be formally located by the dating algorithm. For instance, this was the case for the United States, for which the CLI peak was announced in November 2007, two months before its formal identification in January 2008. This was also the case for the OECD as a whole (figure 19.1). This potential gain, corresponding to distance b in the chronology of events, is difficult to assess in retrospect.

- The identification of the CLI turning points was generally in advance of the first release of the quarterly GDP (difference between columns C and D), thus providing reliable information to policy makers on the forthcoming evolution of GDP. Moreover, the identification of turning points in GDP based on quarterly national accounts' data also requires some time, corresponding to distance e in the chronology of events. While this additional delay is not reported in table 19.1 and table 19.2, it can be estimated to roughly 5 months based on currently available national accounts.

Table 19.1: Identification of the 2007/2008 peak

Countries	CLI Date of peak (based on currently available data)	GDP Date of trough and ex-post lead (number of months)	CLI Date when the peak was first formally identified in real time	GDP Release date (quarter corresponding to the trough)
	A	**B**	**C**	**D**
Canada	Jun-07	Jun-08 [12]	Feb-08	Aug-08
France	Jun-07	Jan-08 [7]	Jan-08	May-08
Germany	Jan-07	Mar-08 [14]	[a]	May-08
Italy	May-07	Feb-08 [9]	Oct-07	May-08
Japan	Jan-07	Feb-08 [13]	[b]	Aug-08
United Kingdom	Jun-07	Jan-08 [7]	Jan-08	Apr-08
United States	Jun-07	Oct-07 [4]	Jan-08	Jan-08

Source: Main Economic Indicators, OECD.

[a]In December 2006, our routine tentatively detected a peak in May 2006 for Germany. The signal remained stable until October 2008 when the same turning point was gradually shifted ahead to reach May 2007. Subsequent revisions have placed that peak in the interval January-May 2007.

[b]A downturn in the CLI for Japan in January 2006 was first identified in September 2006. Since then, this turning point has been shifted to January 2007 as indicated in table 19.1 (first column). However, it would be misleading to consider that the January 2007 turning point was announced as early as September 2006.

19.4 Conclusions

The conclusions that can be derived from this real-time performance analysis were able to anticipate the Great Recession in G7 countries at an early stage, although, by their very nature, OECD CLIs could not give an indication of the depth of the coming crisis. Such results confirm the conclusion previously reached by Gyomai and Guidetti (2011) who relied on an ex-post analysis. Admittedly, the leading properties of

Table 19.2: Identification of the 2009 trough

Countries	CLI Date of trough (based on currently available data)	GDP Date of trough and ex-post lead (number of months)	CLI Date when the peak was first formally identified in real time	GDP Release date (quarter corresponding to the trough)
	A	**B**	**C**	**D**
Canada	Feb-09	Jun-09	Sep-09	Aug-09
France	Feb-09	Jun-09	Jul-09	Sep-09
Germany	Feb-09	Jun-09	Sep-09	Aug-09
Italy	Mar-09	May-09	Jul-09	Oct-09
Japan	Mar-09	Apr-09	Oct-09	Sep-09
United Kingdom	Jan-09	Jun-09	Aug-09	Sep-09
United States	Mar-09	May-09	Sep-09	Sep-09

Note: The difference between columns A and B corresponds to the ex-post lead of the CLI, i.e. to distance $(a + c)$ in the chronology of events (figure 19.2). The difference between columns A and C corresponds to the time required by the dating algorithm to formally identify a turning point, i.e. to distance (a) in the chronology of events. The difference between columns A and D corresponds to distance $(a + c + d)$ in the chronology of events.
Source: Main Economic Indicators, OECD.

the OECD CLIs are less satisfactory if real-time constraints are taken into account. However, statistical and methodological revisions that occurred since the crisis do not seem to have shifted CLI turning points to earlier dates, or to have artificially improved the CLI performance.

The main reason why real-time data are less favourable to the performance of the OECD CLIs is that some time is needed to formally identify turning points once they manifest. About six months are typically required for the formal identification of turning points. This means that, during the Great Recession, CLI turning points could not always be formally identified before the actual occurrence of turning points in GDP. Nevertheless, informal identification based on the sign of CLI growth rates allowed the OECD to announce CLI turning points before they could be formally located by the dating algorithm in some cases. Moreover, most CLI turning points could be identified before national accounts were released and, even more so, before turning points in GDP could be formally identified based on national accounts' data.

Analysing the forecasting errors made by the OECD during the Great Recession in G7 countries, Pain et al. (2014) show that the information provided by the real-time CLIs could have helped to better identify the early stages of the recovery in 2009 (Appendix 7 in Pain et al. (2014)). Economists and statisticians at the OECD are currently working together in order to better exploit the joint potential of the CLIs and the short-term forecasting models.

Annex

19.A Annex

This annex illustrates the method used to assign a score to press release headlines (figure 19.1). Selected keywords in monthly Press Releases have been translated into scores. The selection of such keywords has been made so that the core message of the CLI signal could be fully captured. A positive sign has been assigned to messages identifying a recovery phase and a negative sign to messages identifying a slowdown. Scores are reported in table 19.A.1 below.

Table 19.A.1: CLI Press Releases: Keywords and associated scores

| | **CLI Press release s key words** | |
Left side (entering crisis)	Assigned scores	Right side (exiting crisis)
Improved / continued positive outlook	4	——
Mixed outlook	3	Stronger signals of expansion
Moderating outlook	2	Stronger signs of recovery
Weakening outlook	1	Broad economic recovery
Downswing	0	Signs of improvement
Slowdown	-1	——
Intensified slowdown	-2	Easing pace of deterioration
Sharper slowdown	-3	Deep slowdown but the pace of deterioration is easing
Deepening slowdown	-4	Deep slowdown
Lowest levels since 1970s	-5	——
New low	-6	——

Note: The sign "——" indicates that there is no message associated with the score.
Source: CLI Press releases, OECD.

Bibliography

Astolfi, R., M. Gamba, E. Guidetti and P. A. Pionnier (2016): "The use of short-term indicators and survey data for predicting turning points in economic activity: A performance analysis of the OECD system of CLIs during the Great Recession", OECD Statistics Working Papers, 2016/08, OECD Publishing, Paris, available online at: http://dx.doi.org/10.1787/5jlz4gs2pkhf-en.

Bry G. and C. Boschan (1971): Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. NBER Technical Paper Number 20.

Diebold F.X. and G.D. Rudebusch (1991): Forecasting Output with the Composite Leading Index: A Real-Time Analysis. Journal of the American Statistical Association, Vol.86, Number 415, pp. 603-610.

Fulop G. and G. Gyomai, (2012): Transition of the OECD CLI system to a GDP-based business cycle target. OECD background note, available online at: http://www.oecd.org/std/leading-indicators/49985449.pdf and http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/12/Fich/dt1240e.pdf

Gyomai G., Ahmad N. and R. Astolfi (2017): The OECD System of Composite leading indicators Chapter 8. In Eurostat Handbook on cyclical composite indicators, Luxembourg (2017).

Gyomai G. and E. Guidetti (2011): The Performance of the OECD's Composite Leading Indicator during the 2007 Financial Crisis, The Statistics Newsletter - OECD Issue Number 51, available online at: http://www.oecd.org/std/47632702.pdf.

Nilsson R. and G. Gyomai (2011): Cycle Extraction: A comparison of the Phase-Average Trend Method, the Hodrick-Prescott and Christiano-Fitzgerald Filters. OECD Statistics Directorate Working Paper Number 39, available online at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/doc%282011%294&doclanguage=en.

OECD (2012): OECD System of Composite leading indicators, available online at: http://www.oecd.org/std/leading-indicators/41629509.pdf.

Pain N., Lewis C., Dang T-T., Jin Y. and P. Richardson (2014): OECD Forecasts during and after the Financial Crisis: A Post-Mortem. OECD Economics Department Working Paper Number 1107, available online at: http://www.oecd-ilibrary.org/economics/oecd-forecasts-during-and-after-the-financial-crisis_5jz73l1qw1s1-en.

20 Overview of Existing Business Cycle Clock Applications

Authors

Jacques Anas
Independent Exper
jacques.anas@free.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Rosa Ruggeri Cannata
European Commission, Eurostat
rosa.ruggeri-cannata@ec.europa.eu

Dan A. Rieser
European Commission, Eurostat
dan.rieser@ec.europa.eu

Claudio Parracho-Soares
European Commission, Eurostat
claudio.parracho-soares@ec.europa.eu

Table of Contents

20.1 Introduction

The use of Internet dramatically changed the way in which users expect to access, view and use statistical information. In particular with a large offer of data and indicators, users need to visualize statistical information in a comprehensive manner, quickly and easily.

In this respect, several institutions started in the first decade of XXIth century to develop visualisation tools to facilitate the analysis of business cycle. Eurostat also was very active in this context and developed a business cycle clock (BCC) which was, at least in its first version released in 2005, based on the same methodology of the business cycle tracer of the CBS. The Eurostat BCC's purpose was to portray the up and downswings of euro area, European Union and national economies, displaying the developments of Principal European Economic Indicators (PEEIs). In 2009 Eurostat released a new version with advanced functionalities.

Generally, a Business Cycle Clock is designed to better visualize business cycles and fluctuations of the economic activity. Its great advantage is that it enables to visualize each indicator and its evolution over time; but it is also possible to portray all indicators moving like a "cloud" and thus observing dynamic behaviours and relationships that come into being in a given economic momentum.

The key feature of a business cycle clock is the capacity to represent the development of the overall economic activity and the associated economic dynamics in a timely and understandable way. This means that the business cycle clock must entail robust and reliable indicators (i.e. delivering clear signals without false alarms); ensure freshness and timeliness of the information; and present them in an attractive and easy-to-read format.

The BCC opens to a number of statistical challenges, such as advanced facilities for the analysis of short-term development and the benchmark with cyclical movements; the support for a composite index of economic activities; identification of turning points etc. Furthermore, various IT challenges arise, such as the design of an attractive presentation capable to summarize information at a glance and its dissemination via mobile devices.

At the same time, the beginning of 2000's has been characterised by a growing interest in the development of business cycle composite indicators providing clear signals on the cyclical movements and on the occurrence of turning points. Such indicators had been built up with the aim of estimating in real-time, or anticipating, the main cyclical movements of a given economy, and detect or forecast the turning points.

Unfortunately, the reading of such composite indicators can sometimes be challenging, especially for non-expert users, as the interpretation of the messages delivered by them is often complicated without specific statistical and economic knowledge. In this context, tools like charts and graphs can provide a useful way to present the outcomes of business cycle composite indicators, but not necessarily they do facilitate their readings. Combining graphical tools such as the Business Cycle Clock with the outcome of business cycle composite indicators can provide an easily understandable and fully informative way to display the messages returned by such indicators.

In this chapter, we first describe main characteristics and methodology of existing business cycle, identifying some advantages and drawbacks. In a second stage, we are presenting the new Eurostat business cycle clock based on the historical dating and monthly cyclical indicators presented by Anas et al. (Chapter 14). The description of the new tool is based on Mazzi (2015) and Anas and Mazzi (2015).

This chapter is structured as follows: Section 20.2 presents the main general features of the various business cycle clocks. Section 20.3 details and compares existing BCCs, including the former Eurostat BCC. Section 20.4 presents the main remarks done to Eurostat BCC and proposals for its improvements. Section 20.5 briefly outlines the Eurostat system for cyclical analysis. Section 20.6 and 20.7 respectively describe the main design of the new Eurostat Business Cycle Clock and its structure and interpretation while in section 20.8 two examples related to the practical use of the clock are provided. Finally, section 20.9 concludes.

20.2 Main features of existing Business Cycle Clocks (BCC)

This section introduces the main features generally common to existing business cycle clocks, focusing on methodological aspects, selected indicators and graphical representations.

20.2.1 Historical overview

In 1993, the IFO (Institute for Economic Research, University of Munich) created the first business cycle clock. It was based on two components of the business climate: business situation and business expectations.

It has been followed, in 2005, by the Business Cycle Tracer (BCT) of Netherlands Statistics which proposed a similar framework displaying the cyclical components of selected indicators. Since then, the BCT, also known as Business Cycle Clock (BCC), has been adopted by several countries and institutions to monitor their economy. BCCs have been developed and made available on websites by the OECD and Eurostat as well as by the statistical institutes of Germany, Korea, Malaysia, Brazil, etc.[1]

Moreover, in parallel to these developments, an economic climate tracer was proposed by Gayer (European Commission) limited to survey series.[2]

20.2.2 Methodology

Business cycle clocks generally provide a four-quadrant visualisation for the four distinct stages of the economic cycle. BCCs are based on the representation of the growth (or deviation) cycle. The trend adjustment necessary for the extraction of the cyclical component is performed with statistical filtering procedures (except for survey series which are generally without trend). Peaks and troughs of the growth cycle generally divide one cycle into two phases: an ascendant phase (phase of expansion from trough to peak) and a descendent one (phase of contraction from peak to trough). Each phase is further divided into two sub-phases. As a consequence, the evolution of one cycle is divided into four consecutive sequences which can be represented in a four-quadrant graph.

The descendant phase is divided into a "downturn" step when the deviation to trend (or output gap) remains positive and a "slowdown" (sometimes misleadingly worded as "recession") step when the deviation to trend becomes negative. The term "recession" often used in the business clock is confusing because it does not correspond to the classical definition of a recession. The basic definition was given by Burns and Mitchell in 1946: "Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic." We also can refer to the current NBER definition of a recession: "A recession is a persistent period of decline in total output, income, employment and trade, usually lasting from six months to a year, and marked by widespread contractions in many sectors of the economy."

Similarly the ascendant phase is split into an "upswing" (or "recovery") step and a "boom" (or "expansion") when the deviation-to-trend turns positive. Therefore, the indicators shown in the clock are presented in their trend-adjusted form (generally in level on the y-axis and in variation on the x-axis).

---

[1] See section 20.7 for references to different Business Cycles Clocks.

[2] See Gayer, C. (2010) "Report: The Economic Climate Tracer - A tool to visualise the cyclical stance of the economy using survey data", available online at http://www.oecd.org/dataoecd/12/47/39578745.pdf.

20.2.3 Indicators and graphical representation

Any quantitative or qualitative series may be used in a BCC. They have to be de-trended and smoothed out. Various criteria may be used for the selection. The number of indicators varies from four in the OECD BCC to fifteen in the CBS BCT.

The y-axis represents the deviation to trend of the series while the x-axis represents the growth rates of the deviation to trend. Growth rates may be monthly, quarterly or semi-annual growth rates.

Most clocks aim to represent the cycle of reference series (GDP or industrial production index) and basic indicators. For instance, the OECD clock shows their composite leading indicator (CLI), and the former Eurostat clock proposed fifteen Principal European Economic Indicators.

The first methodological step is to de-trend the series. The cyclical component is usually extracted using the Christiano-Fitzgerald (CF) or Hodrick-Prescott (HP) band-pass filters. The series are then normalized to have an average of zero and a unit standard deviation. By this way, they fit within the same graphical frame. The y-axis shows the normalized deviation-to trend of each indicator. The x-axis shows the one-period growth rate of the de-trended series. Representations can include a supplementary graph showing the growth rates of the original series, these must not be confused with the growth rates shown on the usual BCC graphs.

BCCs are usually based on the same rationale: a four-quadrant representation with generally a counter-clock reading of the cyclical evolution. The aim is to give a dynamic reading of the cyclical evolution, either showing the various indicators for one country, or one single indicator for several countries, or a combination of both.

20.3 Overview and comparison of existing Business Cycle Clocks

In this section, we first present some business cycle clocks: the IFO Business Cycle Clock, the Business Cycle Tracer of Statistics Netherlands, the OECD Business Cycle Clock, the German Business Cycle Monitoring and finally the former Eurostat Business Cycle Clock, the Mexico BCC and the South Korea one. Next, we compare the mains features of those business cycle clocks.

20.3.1 Overview of existing Business Cycle Clocks

IFO Business Cycle Clock

The IFO Business Climate was developed by the Ifo Institute in the mid-1960s on the basis of the monthly Ifo Business Survey (see Abberger and Nierhaus (2007)). The business climate indicator is determined according to the formula $[(BS + 200)(BE + 200)]1/2 - 200$, where BS and BE designate the percentage responses balance regarding the current business situation and the business outlook in the next six months. The IFO Business Climate was published for the first time in 1971, for manufacturing only. One year later, the business climate data for the survey sectors manufacturing, construction, wholesaling and retailing were combined to form the indicator in its present form (Ifo Business Climate for German Industry and Trade).

The four-quadrant scheme for the cyclical relationship of business situation and business expectations from the Ifo Business Survey (IFO BCC) was published for the first time in 1993. Initially the variables in the economic cycle moved counter clockwise due to the arrangement of the axes (see Leibfritz and Nierhaus, 1993). The present quadratic depiction of the Ifo Business Cycle Clock - in which the direction of the business situation and expectations move in the conventional clockwise direction - was introduced in 1999 (see Abberger

and Nierhaus (2010)).[3]

Statistics Netherland's Business Cycle Tracer

The basic idea of the business cycle tracer originated from Gert Buiton from CBS Statistics Netherlands. His goal was to develop a representation of short-term economic indicators which would enable a quick analysis of the state of the business cycle, via a disaggregated multivariate approach. He aimed at analysing the current state of the business cycle. Methodological aspects have been developed in 2005 by Floris van Ruth, Barry Schouten and Roberto Wekker.[4]

According to Statistics Netherlands, the Business Cycle Tracer is a system which acts as a coincident indicator of the Dutch business cycle (understood as "growth cycle"). It maps real-time business cycle developments by tracking the cyclical development of a selected set of lagging, coincident and leading economic indicators. The disaggregated approach enables a detailed analysis of the state of the economy.

The choice to use the growth cycle is motivated by the fact that very different economic regimes are in place when an economy is above or below potential, each one with its own characteristic developments. If an economy is exhibiting positive growth but is also growing (far) below its potential, phenomena associated with recession may still manifest.

The cycle extraction is done in various steps. First the series are seasonally adjusted and smoothed out by the Census X12 program (use of the Henderson filter). Second, the cycle is extracted with a Hodrick-Prescott filter (the cut-off parameter differs among indicators) and finally normalized.

Statistics Netherlands first launched its Business Cycle Tracer at the end of 2005. Every month, the fifteen selected indicators combine to show in one snapshot whether the economy is performing "well" or "poorly" (above or below the long-term trend) and whether it is "improving" or "deteriorating" (increase or decrease compared with the previously measured period). By showing these snapshots consecutively, the business cycle evolution is displayed. The horizontal axis represents the period-on-period change of an indicator and the vertical axis represents its distance to trend. Another way of interpreting this set-up is that the vertical axis gives the state of the indicator, and the horizontal axis its direction of change. Recently, a new column has been added - on the right-hand side of the Tracer, next to the indicator selection column - making it possible to display a trail for each indicator. The trail shows how that indicator developed in previous periods. The span of periods - and thus the trail - can be set between one month and five years.

The BCT allows to plot a maximum of fifteen indicators carefully selected: the indicators are usually classified as being leading, lagging or coincident with respect to the general business cycle, and are used to construct, leading, lagging and coincident aggregate indicators. The approach of Statistics Netherlands aims to construct a system showing at a glance all important aspects of the business cycle. This means combining leading, coincident and lagging indicators in one system. By using a balanced mix, the whole set was made on average to be coincident with the cycle.

A first strict selection process among forty potential indicators was based on various criteria and particularly: maximum correlation (including lag) of its cycle with the GDP cycle, whether and how fast the indicators detected the major turning points in 1990, 1994 and 2000, the check for the presence of idiosyncratic or minor cycles. The resulting twenty selected indicators were then assessed based on their real-time turning point detection performance. In practice, it was shown that the lead of most indicators was clearly reduced, sometimes disappearing altogether. On average, a two to three month delay is apparently introduced when evaluating real-time performances.

One important consequence of this is that a system to track the business cycle has to be slightly leading ex-post to shoe a coincident behaviour in real-time. Finally fifteen indicators were retained: GDP, producer

---

[3]The "IFO Business Cycle Clock: Circular correlation with the real GDP." CESifo working paper number 3179.

[4]The Statistics Netherlands' Business Cycle Tracer. Methodological aspects; concept, cycle computation and indicator selection, Floris van Ruth, Barry Schouten and Roberto Wekker, Statistics Netherlands woking paper 2005-MIC-44.

confidence, unemployment, consumer confidence, employment, temporary jobs, purchases of Durables by consumers, exports, fixed capital formation, business orders received, household Consumption, industrial production, vacancies, 10-year bond yield and bankruptcies.

Denmark: Business Cycle Tracer

Statistics Denmark also disseminates a Business Cycle Tracer. The business cycle is based on the tendency survey of the businesses of manufacturing, construction, service and retail trade, conducted by Statistics Denmark. Based on those data, a composite indicator is calculated by Principal Component Analysis for each sector. More information is available on Statistics Denmark website.[5]

Contrary to a traditional graphical representation of the economic cycle on a timeline, data in a business cycle have been traced in a system of coordinates defined by the four economic stages: increase above trend, decrease above trend, decrease below trend and increase below trend. The four stages are also called; boom, downswing, recession and upswing.

The visual course enables one to compare a current economic cycle with earlier economic cycles, in regard to how fast the change occurs and the size of the fluctuations compared to underlying trends. The course is subject to a certain degree of uncertainty especially when looking at the latest month's development, which is continuously revised, but overall, the results are sturdy.

The economic cycle is a graphical presentation of the market trend development created by a method similar to what is used by the EU Commission (DG ECFIN). The interactive presentation tool has been provided by Statistics Netherlands. The tool can be found at: Business Cycle Tracer.[6]

The OECD business cycle

The OECD Business Cycle Clock has been designed in 2008, based on the original concept of Statistics Netherlands (Business Cycle Tracer). It aims at better visualizing business cycles (fluctuations of economic activity around their long term potential level) and how some key economic indicators interact with the business cycle. The Clock was put on line on the OECD website in 2009, without major changes. The tool aims at showing the leading, coincident or lagging behaviour of three composite indicators: the OECD CLI (composite leading indicator) and the (smoothed) standardised business and consumer confidence indicators. The reference series has been changed from the industrial production index to GDP in 2012.[7]

The dynamic graph enables to plot these four indicators simultaneously, moving in anticlockwise direction through the four quadrants over time. The x-axis represents the annualised 6-months growth rate and the y-axis represents the level of the series. Because of the phase shift introduced by the 6-months growth rate, peaks and troughs occur a few months before the indicators cross the y-axis.

The four indicators are:

- the amplitude-adjusted CLI (so that its typical variation is of the same order of magnitude as variations in the reference series) and the de-trended GDP (a double HP filter, keeping frequencies within 2 years and 12 years).

- The standardised business (restricted to manufacturing) and consumer confidence indicators which are smoothed using a HP filter removing shorter than 6 months cycles, and thereafter normalised and

[5]Statistics Denmark: http://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/business-cycle-tracer.
[6]Statistics Denmark: http://www.dst.dk/en/Statistik/emner/erhvervslivet-paa-tvaers/konjunkturbarometre/konjunkturcyklus.
[7]An OECD study has demonstrated that the IPI was not an ideal coincident proxy of the GDP anymore. Although the IPI works well as a GDP proxy for some countries, it was shown that for the majority of the OECD economies and major non-member economies the cyclical components of the IPI and GDP were not sufficiently synchronized. Therefore, in April 2012, the IPI proxy was phased out from the OECD CLI system and the reference cycle is now estimated directly from GDP figures except for China.

amplitude adjusted.[8]

Data are available for all OECD countries (including regional aggregates, in particular the euro area) and other major economies (Brazil, China, India, Indonesia, Russian Federation, South Africa.). The data set goes back to 1970. A maximum of two countries or areas may be plotted together. Each indicator can be shown with a trail of 24 months maximum.

Classical options for the graphs are available: play speed, starting date, selection of series and countries and trail length. The wording of the two different sub-phases of the descendant phase is varying. On the website, the two distinct quadrants of the descendant phase are labelled as:

- Slowdown: series decreasing but above 100.

- Recession: series decreasing and below 100.

According to an internal OECD document , "in our [i.e. the OECD's] CLI terminology, the contraction phase is further divided into two: the downturn step with a CLI decreasing but above 100 and the slowdown step meaning that CLI is decreasing and below 100". As a consequence, there is confusion about the definition of a recession.[9]

Destatis business cycle monitor

The quarterly Business Cycle Monitor (BCM) for the Real GDP was developed by the German Federal Statistical Office (DESTATIS). It was a new monitoring system aiming at depicting the developments of the global economic activity in a four quadrant scheme.[10] The Federal Statistical Office has used the Netherlands Central Bureau of Statistics approach in developing its Business Cycle Monitor. The BV4.1 time series analysis procedure was used instead of Census X12 to remove noise and seasonality.

For the selection of series, the Federal Statistical Office followed the guiding principle of the Netherlands Central Bureau of Statistics to include a set of data of indicators that lead or lag behind in relation to the short-term economic trend. The current version of the monitor comprises series of indices of new orders, production and turnover for manufacturing and partly also for selected main groupings.

The BCM is available on line, based on 13 series. It shows a dynamic evolution since 1993, but without any possible trail. The reading is clock-wise since the typical axes have been exchanged: the abscissa shows the position respect to trend, while the ordinate shows the variation of the de-trended series.

The former EUROSTAT Business Cycle Clock

The former Eurostat BCC was based on the Statistics Netherlands Business Cycle Tracer. It can plot a maximum of 15 indicators at the same time: GDP, consumption, investment, export and import, inflation, unemployment, labour costs, employment, industrial production, industrial producer prices, production in construction, deflated retail sales, deflated turnover, industrial new orders and Economic Sentiment. Except for ESI, series are generally available from 1990 (some from 1995).

The graphical layput of the former Eurostat BCC was also based on the Statistics Netherlands' one. Its first website version was launched in July 2008. In April 2009 Eurostat introduced a new version of the BCC with advanced functionalities.[11] The Eurostat BCC used the classical four-quadrant representation but no wording was made for each of the four stages of the economic cycle making its interpretation more difficult.

---

[8]The series are normalised by subtracting their mean and then dividing the difference by their standard deviation. After normalisation, they are amplitude-adjusted to the de-trended indices of GDP and finally centred around 100.

[9]"The OECD Business Cycle Clock", Emmanuelle Guidetti, September 2012.

[10]See: Oltmanns, E. (2009), "Das Bruttoinlandsprodukt im Konjunturzyklus", Wirtschaft und Statistik 10/2009, 963-69.

[11]Among these new functionalities, it is worth mentioning the switch indicators/countries option and the possibility to download the graphical animation for running it off-line, i.e. without a connection to the Internet.

This version of the Eurostat BCC was abandoned in 2015 due to data availability problems and also because Eurostat was planning to develop an enhanced version of the BCC.

The Mexico business cycle clock developed by INEGI

The Mexican Business Cycle Clock shows in a graphic way where the general economy and some selected economic aggregates are located in the business cycle. The MBCC was developed by INEGI during the second half of 2010 based on the original Business Cycle Tracer. The MBCC is based on a set of sixteen indicators: the INEGI coincident indicator and its components (monthly GDP; Industrial production index; Retail sales; Workers registered at the Mexican Institute of Social Security; Urban unemployment rate and Total imports); the INEGI leading indicator and its components (the price index of the Mexican stock exchange in real terms; the real exchange rate; the interest rate; the Standard & Poor's 500 Index (U.S.); the manufacturing employment trend; and Non-oil exports); plus the producer confidence and consumer confidence indicators.

As in the case of other clocks, the MBCC interface has a main clock with animation control buttons and a table to select indicators. Up to six of the sixteen available indicators can be shown simultaneously beside the MBCC. The combination of the main clock and the set of graphics show how the cyclical component of the indicators moves through the quadrants while locating their position relative to their long-term trend. The typical behaviour of the cyclical component of an indicator on the four quadrants of the clock is counter-clockwise.

The Statistics South Korea business cycle clock

In 2009 the Economic Statistics bureau together with Statistics Information Bureau of South Korea developed the BCC. It was based on the same methodology of the BCT, using the following monthly series: Industrial Production Index, Service Industry Activity Index, Consumer Goods sales Index, Index of Equipment Investment, Value of Construction Completed (Real), Exports (Real), Imports (Real), Number of Employed Persons, Business survey Index (manufacturing) and the Consumer Expectation Index.

The selected variables constitute a balanced set of indicators, chosen in a way to obtain a coincident picture of the economic cyclical situation. Selected variables are detrended by a double-sided HP filter. The BCC has been designed as a complement to the South Korean system of Composite Leading, Coincident, and Lagging Indexes.

20.3.2 Comparison of existing Business Cycle Clocks

The methodology at the base of BCCs is quite similar among the applications described in subsection 20.3.1, generally an indicators selection procedure is followed by a detrending step and a representation one. As we have seen filters can vary form one tool to another.

The main difference will then be the selected indicators, which could be tailored to specific characteristics of the economy. Indicators can be basic variables, or in some cases confidence indicators or composite ones. OECD for example includes in the clock the CLI which is expected to provide more meaningful information. The same applies to Mexico which includes both the coincident and leading cyclical indicator while Denmark is including a composite indicator derived from business and consumer surveys. Set of indicators are usually balanced, that is indicators which are leading, coincident or lagging with respect to the business cycle are chosen in a way to produce a set of indicators which will exhibit a coincident behaviour with the development of the economic situation. It is important to perform this kind of assessment in real-time, because results can vary and lag effects could appear. Another consideration to keep in mind when looking at indicators selected for a BCC is their ability of representing all major sectors of an economy. Other practical limitations could be the availability at high frequency of the indicators and the length of the available time series.

Finally, in term of visual representation, all the analysed BCCs display a four-quadrant graph. Several BCCs present the trail of the indicators, i.e. display the past positions of an indicator, in order to better follow its evolution, including the former Eurostat BCC. In table 20.1 below we are comparing the main features of the BCCs, as analysed in detail in subsection 20.3.1.

Table 20.1: Comparison of business cycle clocks

	Eurostat	Statistics Netherlands	OECD	Statistics Denmark	DESTATIS	INEGI Mexico	Statistics South Korea
Name	Business Cycle Clock	Business Cycle Tracer	Business Cycle Clock	Business Cycle Tracer	Business Cycle Monitor	Business Cycle Clock	Business Cycle Clock
Indicators	15	15	4	18	13	16	10
Countries	36 + EU15 + EU27 + EA17	Netherlands	34 OECD + 6 Non OECD + aggregates	Denmark	Germany	Mexico	South Korea
Filter	CF	HP	HP	HP	HP	Two side HP	HP

Source: Own research

20.4 Remarks on the former Eurostat business cycle clock and lines of improvement

In this section, we first review some drawbacks of the former Eurostat BCC. Next, we present the main lines which guided the development of the new Eurostat business cycle clock in 2016.

20.4.1 Remarks on the former Eurostat business cycle clock

Two main remarks may be done on the former Eurostat BCC and could also be extended to other BCCs.

First, it presented the evolution of a number of basic indicators which may at times provide diverging signals on the economy and leave the user perplex about the final message. Thus, to provide a clear message on the economic cyclical situation, the representation of few composite cyclical indicators could be preferred to the representation of many basic indicators. This comment also holds for several BCCs; in this area, interesting alternatives are the OECD and the INEGI business cycle clocks. As already said in subsection 20.3.1, the OECD has chosen to complement the BCC by displaying the Composite Leading Indicators too, to send clearer and synthetic signals.

The second remark relates to the impossibility of identifying the "economic recessions" defined according to the Burns and Mitchell definition as outlined above. The "economic recessions" as defined by Burns and Mitchell (1946) cannot be identified by the former Eurostat BCC. Indeed, the descendent phase of the cycle is divided into a "downturn" step when the deviation to trend (or output gap) remains positive and a "slowdown" (sometimes misleadingly worded as "recession") step when the deviation to trend becomes negative. The term "recession" often used in business cycle clocks is confusing because it does not match the classical definition of a recession.[12]

---

[12]"Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic." Burns and Mitchell (1946).

20.4.2 Lines of development for the new Eurostat BCC

In order to overcome the pitfalls listed in the previous section, the following decisions were taken in the design of the new Eurostat BCC (see Anas and Mazzi (2015) and Mazzi (2015)):

1. Using few cyclical composite indicators in order to provide clearer messages about the cyclical situation. The selected cyclical indicators are composite indicators for the turning point detection of the acceleration cycle (ACCI), growth and business cycle (MS-VAR GCCI and MS-VAR BCCI) as described in Anas et al. (chapter 14) and Billio et al. (2016).

2. Looking for a cyclical representation based on Eurostat-compliant ABCD approach and possibly the extended αABβCD approach (Anas et al. (2008) and Anas et al. (chapter 14). This allows for a clear distinction between expansions, slowdowns and recessions.

3. Complementing the real-time recession signals delivered by the turning point indicators with the reliable estimates of past turning points provided by the Eurostat historical dating chronology developed in Anas et al. (2005) and in Anas et al. (chapter 14).

By combining those elements, it has been possible to obtain in a graphically attractive way a comprehensive overview of the cyclical situation. In particular, the historical dating chronology provides a robust look at past cyclical events while the system of turning points detection describes the probability of turning points' occurrence in real-time.

The main idea of the new representation is to display, on a dynamic page, the past turning points, as obtained in the historical dating chronology, in a classical graph showing the evolution of the GDP, while the results of the indicators for turning point detection are presented in a clock-style representation reproducing all phases of cyclical movements.

The historical chronologies and the system of euro area and member states turning point detection are based on quite substantially different principles: in the first one the turning points are identified by a purely non-parametric rule and we are trying to keep them as much as possible invariant in the past. By contrast, the indicators for turning points detection are based on probabilistic rules and are subject to either type I or type II errors. This explains why it is possible that, over the same time period, there is a disagreement between the two. See Anas et al. (chapter 14) for more details.

For the past, the disagreement can mainly be attributed to either type I or type II errors of the probabilistic indicators so that the historical dating is considered as more reliable. For the latest and the current period the situation is quite different, because the probabilistic indicators are usually based on more updated information than the historical dating which by the way stops earlier in time. For this reason, there is a tendency to privilege the message provided by the probabilistic indicators for the very recent and current periods, waiting for a final validation which will be provided by the historical dating later on.

# 20.5 Methodology: Eurostat system for cyclical analysis

In this section we are recalling the main characteristics of the Eurostat system of cyclical information which constitute the engine of the new Eurostat business cycle clock. For a more technical and complete description, refer to Anas and Ferrara (2004), Anas and Ferrara (2004), Anas et al. (2008), Anas et al. (chapter 14), Billio et al. (2016) and Mazzi (2015).

Since the end of 2006 Eurostat is monitoring the euro area cyclical situation by regularly producing two sets of relevant indicators: a quarterly updated historical dating and a monthly updated set of probabilistic turning points coincident indicators. Those two sets of indicators have been firstly developed for the euro area and

then extended progressively to cover all member states. In a first phase they were referring to the growth cycle and to the business cycle, while later also the acceleration cycle has been included in order to monitor a full sequence of turning points as defined by the αABβCD approach (see Anas and Ferrara (2004), Anas and Ferrara (2004) and Anas et al. (2008)). Concerning probabilistic indicators, those for growth and business cycle are based on a MS-VAR representation which jointly returns growth cycle and business cycle recession probabilities (see Billio et al. (2016) and Mazzi (2015)). By contrast, the indicator for the acceleration cycle is based on a univariate Markov-switching model.

The turning point detection issue is considered as a progressive and almost real-time follow-up of the cyclical movement. Even if no cycle is similar to the previous one, the sequence of turning points is always respected in practice. Figure 20.1 below shows the turning points of the extended αABβCD approach.

Figure 20.1: The extended αABβCD approach
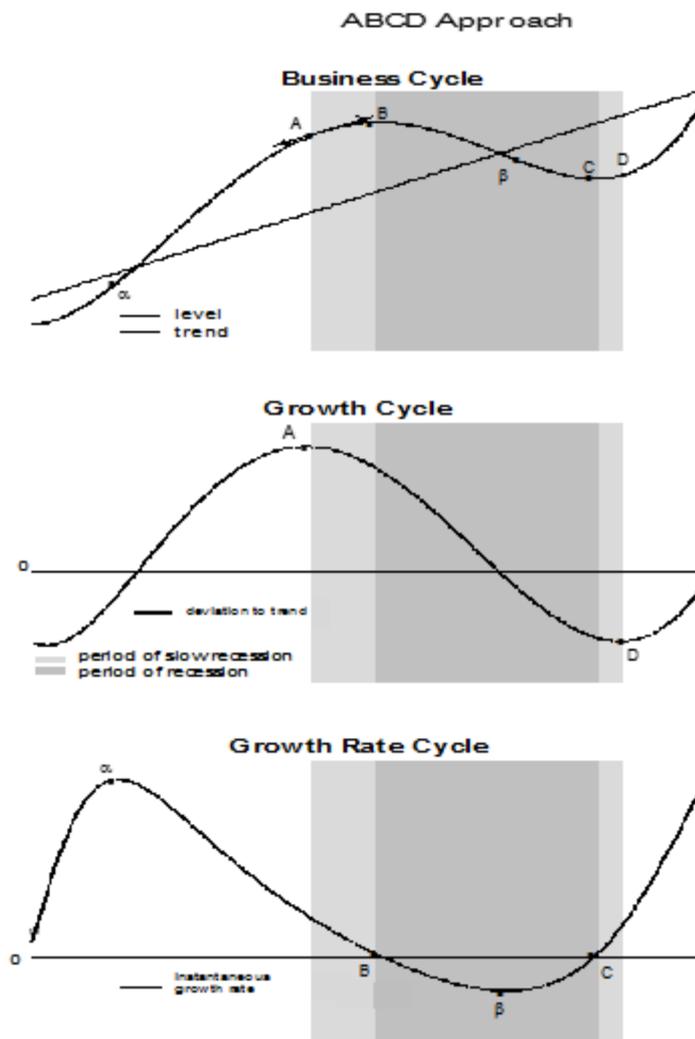
Source: Authors' calculations

Deceleration (point α) will turn into a slowdown movement (point A of the growth cycle) and if it is getting worse, the growth rate will become negative (point B), thus determining a recession. B and C are the extrema of the business cycle; A and D are the extrema of the growth cycle (also called deviation cycle) and α and β are the extrema of the growth rate cycle (also called acceleration cycle).

The Growth Cycle Coincident Indicator (GCCI) provides the probability of a slowdown in the economy. Thus it signals the peaks and troughs of the growth cycle. The Business Cycle Coincident Indicator (BCCI) provides

the probability of a recession signalling the peaks and troughs of the business cycle. Finally, the Acceleration Cycle Coincident Indicator (ACCI) provides the probability of a deceleration in the growth rate and signals the peaks and troughs of the growth rate cycle.

The outcome of those cyclical indicators can be presented either in a graphical or in a tabular form. Figures 20.2 and 20.3 show, for the euro area, the evolution of the univariate acceleration cycle coincident indicator (ACCI) and the multivariate growth cycle and business cycle coincident indicators (MS-VAR GCCI and MS-VAR BCCI), respectively. They also show the results of corresponding historical dating chronologies.

Figure 20.2: Euro Area ACCI univariate
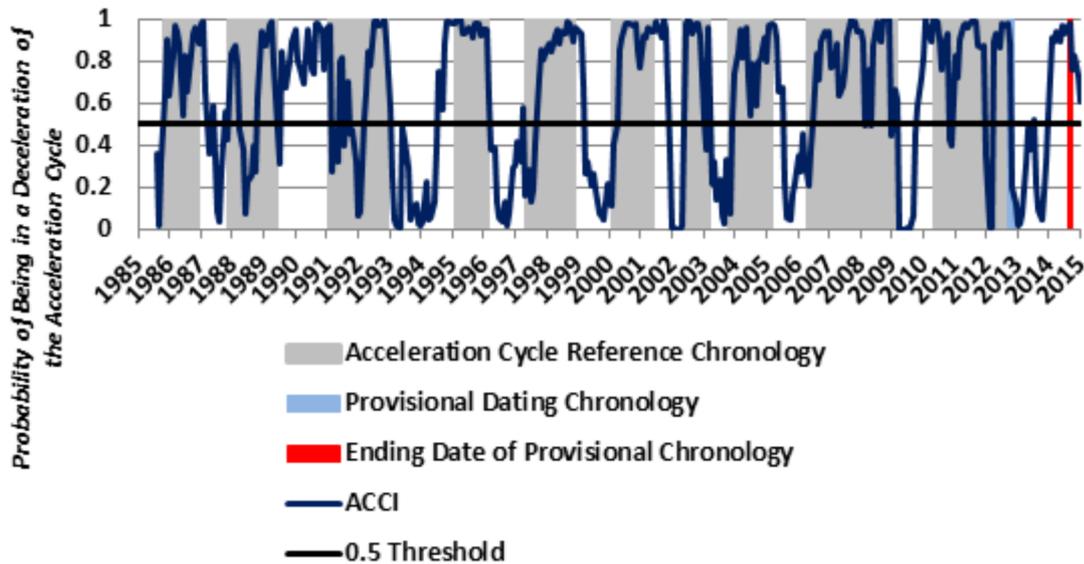
Source: Authors' calculations

Both figures show in the vertical axis the recession probabilities of each indicator and in the horizontal one the time scale. The horizontal line at 0.5 indicates the chosen threshold. Figure 20.3 shows MS-VAR BCCI on the top panel and MS-VAR GCCI in the bottom one since they are simultaneously derived from a single multivariate model. The grey areas in both figures indicate the historical dating chronology.

While the interpretation of the indicators' outcome is not particularly challenging for advanced users, it is quite clear that this is not a friendly way to present the data to a wider audience. Furthermore, since the indicators are presented individually, it is not easy to understand the relations between them so that the global assessment of the cyclical situation, which is the main added value of this system of turning point detection, is often hidden.

In order to overcome this, within the new Eurostat BCC, the outcome of the three probabilistic indicators is presented in a clockwise representation intended to provide an intuitive, easy to read and user-friendly picture of the real-time cyclical situation. In this way the three probabilistic indicators are not directly shown but used as the hidden engine to animate the clock.

The layout, structure and interpretation of the new BCC based on the Eurostat dating chronologies and the Eurostat composite indicators for turning point detection are presented in sections 20.6 and 20.7 respectively.

**Figure 20.3:** Euro Area MS-VAR BCCI and GCCI multivariate
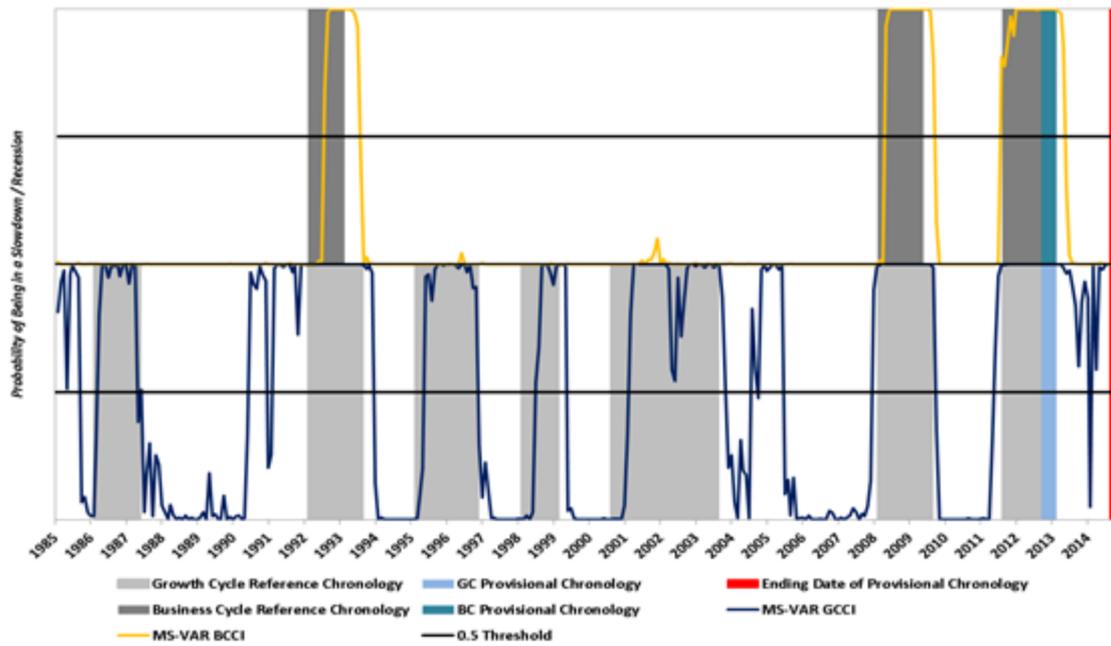
Source: Authors' calculations

20.6 Layout of the new BCC

In the present section we describe the layout of the new Business Cycle Clock, as it is displayed in Eurostat's website.

Figure 20.4: Layout of the new business cycle clock


Source: Authors' calculations

Figure 20.4 is divided in three main parts: in the upper part, there is an historical graphical representation based on the evolution of GDP; in the lower left corner, one or more clocks are displayed with its hand showing the phase at a given time in the euro area and/or member states; while in the lower right corner some statistics associated to the cycles are presented.

The upper part contains a didactical representation consisting of the GDP deviations from the trend, where the peaks and troughs of the cycles are highlighted. The slowdown phases are represented in pink; the recession phases are represented in dark pink; each point of the $\alpha AB\beta CD$ cycle is represented by a vertical line. The graph is based on the data obtained by the historical dating mentioned in the previous section. For this reason it does not contain information for the latest time periods but it provides an historical overview of the cycles over a long time horizon. It is worth noting that the clock and graph representation are dynamic: a play button sets time running. The current position in the graph representation is highlighted and the clock hand runs. The clock on the lower left part is structured according to the $\alpha AB\beta CD$ approach.

20.7 Structure of the clock

This section presents in more detail the structure of the clock itself: the interpretation of the position of the hand as well as the meaning of each sector in the quadrants. Figure 20.5 shows the clock, its turning points and the six sectors.

Figure 20.5: Clock based on the $\alpha AB\beta CD$ approach
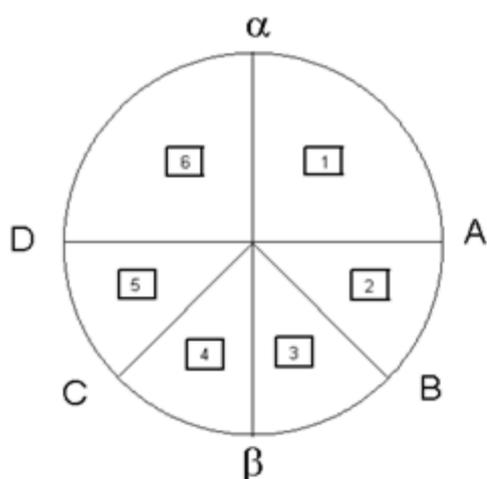
Source: Authors' calculations

The clock is designed so that:

- Noon/midnight is α, peak of the growth rate cycle
- 3 am/pm is A, peak of the growth cycle
- 4.30 am/pm is B, peak of the business cycle
- 6 am/pm is β, trough of the growth rate cycle
- 7.30 am/pm is C, trough of the business cycle
- 9 am/pm is D, trough of the growth cycle

Those turning points delimite six sectors in the clock which correspond to various phases of the business cycle. The location of the hand in the clock is based on the values of the three cyclical coincident indicators for

the acceleration, growth and business cycles described previously, as well as on their positioning with respect to the 0.5 threshold.

The six sectors can be read as follows:

- The upper and lower right quadrants (sectors 1, 2 and 3) show a decrease in the growth rates. In the first sector, the growth rates are still above the trend growth rate but they are decelerating. In the second sector, the growth rates slip below the trend growth rates. In the third sector, the growth rates are negative and at β the growth rate reaches a minimum.

- The lower and upper left quadrants (sectors 4, 5 and 6) show a progressive increase of growth rates. In sector 4 growth rates are still negative but they are accelerating (the size of decreasing is progressively reducing). In sector 5, the growth rates came back to a positive sign but they are still below the trend growth rates and in sector 6 the growth rates are above the trend and progressively accelerating.

Visually, the clock can be read as follows:

- The lower half of the clock (sectors 2, 3, 4 and 5) represents below-the-trend growth rates. The bottom region between B and C, i.e. between 4.30 and 7.30 would depict the recession. It corresponds to one quarter of the graph. The two half quarters of the clock between A and B (3 and 4.30) and between C and D (7.30 and 9) depict the slowdown. They surround the recession area.

- The upper half of the clock (sectors 6 and 1) represent the above-the-trend growth rates (expansion growth cycle phase).

With this representation, the arrow could move forward and backward. It would also enable the visualization of pure acceleration cycles (jump from sector 1 to sector 6) and pure growth cycles (jump from sector 2 to sector 5).This clock representation of the cyclical situation allows for a static and dynamic analysis of the euro area and each member country, as well as for cross countries comparisons.

By using the growth, business and acceleration cycles indicators we compute the location of the hand, associating each clock sector to values of the indicators. Table 20.1 synthetically presents this relationship with the cyclical indicators.

Table 20.1: The clock sectors and the cyclical composite indicators

|		ACCI			
		<0.5		>0.5	
		BCCI		BCCI	
		<0.5	>0.5	<0.5	>0.5
GCCI	<0.5	6		1	
	>0.5	5	4	2	3

Source: Own research

From table 20.1, we can say that in sector 1 the economy is growing above the trend but its growth is progressively decelerating. In sector 2, the still positive growth is below the trend while in section 3 the growth becomes negative. In sector 4, the negative growth starts to accelerate approaching zero. In sector 5, the growth becomes positive but still below the trend, while in sector 6 the economy is growing above the trend and accelerating.

20.8 Examples

In this section we are replicating the examples presented in Mazzi (2015). The two examples presented here will try to answer, with the help of the new Eurostat business cycle clock, two key questions for which the analysts would like to have a clear answer. The first question is related to the identification of which economies are growing above the trend and which ones are still below. Since the growth cycle is defined as the deviation from the trend, so that $GC_t = Y_t - Y_t^p$ for $t = 1, 2, T$; where GC is the growth cycle, Y is the actual growth and $Y^{(}p)$ is the trend, it is possible to show that the growth cycle will cross the trend in A (in a descending phase) and in D (in an ascending phase) of the clock. By drawing a line between A and D we can conclude that, in the sectors of the clock above the line, the economy is growing above trend (sectors 6 and 1), while in the others the economy is growing either below trend or even decreasing. This is shown in figure 20.6.

Figure 20.6: The clock and the economic growth
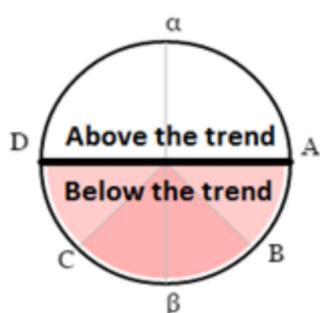
Source: Authors' calculations

The main difference between staying in sector 6 or in sector 1 is that, in the first case, the economy is growing above the trend and it is still accelerating, while in the sector 1 it has started a deceleration phase while still growing above the trend. By using those results we can analyse, in a comparative way, the growth of some Euro area member countries by referring to figure 20.7.

By looking in detail to the various clocks it emerges that almost all economies are growing above trend except Belgium which is still growing below. Furthermore, the Euro area is still in an acceleration phase, while Spain has achieved the peak of the acceleration cycle. For the remaining economies, we can observe that Italy is just at the beginning of the deceleration phase while France, Germany and the Netherlands, as well as Portugal, have a more consolidated deceleration phase. Since for all those economies, the hand is located in the first half of the sector 1, we can also conclude that the risk of reaching point A and therefore starting to grow below the trend, is very low. In this case, by combining the information delivered by the dashboard and the one delivered by the clock, it is possible not only to rank the countries according to the intensity of growth (above or below the trend) but also to obtain useful insight related to the acceleration or deceleration of their growth.

The question concerns the degree of cyclical synchronisation among Euro area countries. To answer the question, we analyse the evolution of the cyclical situation, represented by a series of clocks at different points in time.

In figure 20.8, we consider the Euro area plus six member countries (Germany, France, Italy, Spain, the Netherlands and Belgium). Their cyclical situation is assessed respectively in June 2011, 2012 and 2013. By looking at the countries and Euro area behaviour at the three different points in time, we can observe the following:

- June 2011 shows the immediate entering into slowdown of Spain, Belgium and the Euro area, while Germany, France and Italy are still in expansion; the Netherlands is the only country already in reces-

Figure 20.7: Cyclical situation of the Euro area and some member countries in June 2015
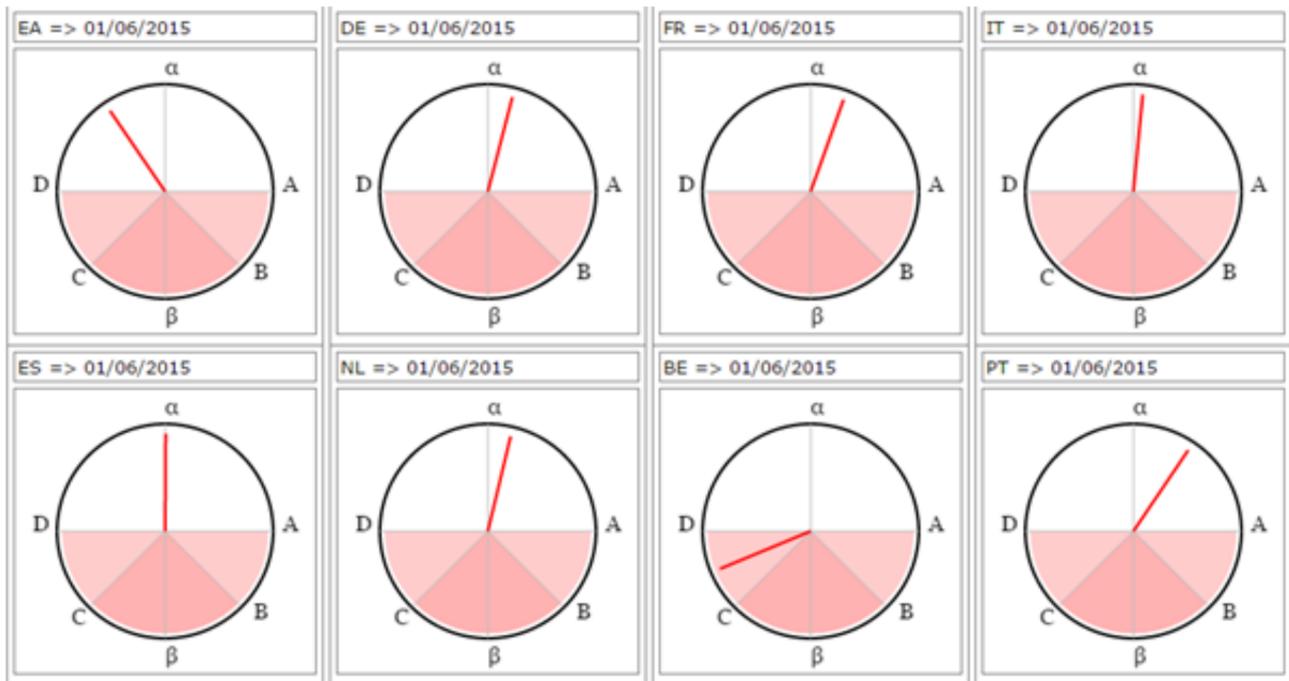
Source: Authors' calculations

Figure 20.8: Analysis of the 2012/2013 recession at Euro area and member countries level
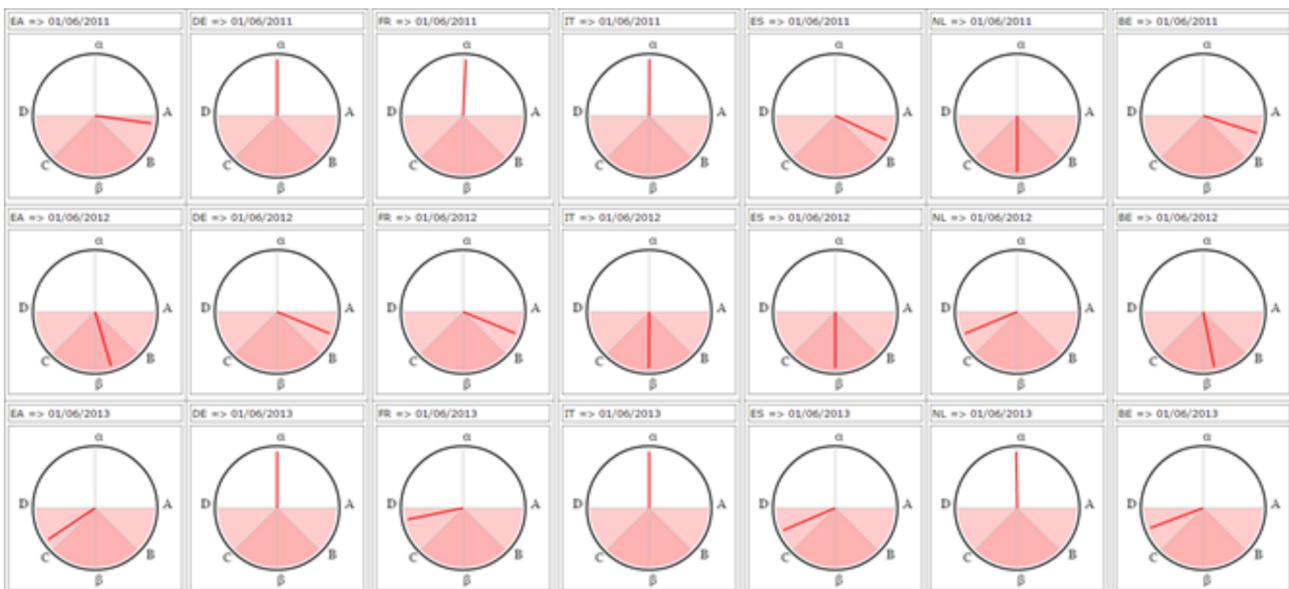
Source: Authors' calculations

sion, achieving the trough of the acceleration cycle.

- June 2012 shows Spain, Belgium, Italy and Euro area in recession. Germany and France are in slow-down only. The Netherlands exited the recession but remained in slowdown.

- June 2013 shows the exit of Euro area, Spain and Belgium from recession, remaining in slowdown, while France continues to be in slowdown. Germany, Italy and the Netherlands are in expansion.

20.9 Conclusions

After having presented an assessment of some already existing business cycle clocks we have introduced the new Eurostat Business Cycle Clock and its main characteristics. The new Eurostat Business Cycle Clock is a dynamic graphical approach for the presentation of a complex model aiming, among others, to facilitate data understanding and interpretation, especially for non-expert users.

The new BCC presented in this paper constitutes, in our view, a relevant step forward with respect to the previous one. It is a real-time monitoring tool for the cyclical situation of the euro area and its member countries, allowing for both temporal and cross-country analysis and comparison. It also enables to go deeper in the details of all phases of the cycles due to the joint analysis based on acceleration, growth and business cycles.

The new business cycle clock presented in this paper is based on the probabilistic turning point indicators elaborated each month by Eurostat as well as on the historical dating chronologies compiled on quarterly basis.

Bibliography

Abberger, K. und W. Nierhaus (2007), "Das ifo Geschäftsklima: Ein zuverlässiger Frühindikator der Konjunktur", ifo Schnelldienst 60(5), 25-30.

Abberger, K. und W. Nierhaus (2010), The "IFO Business Cycle Clock: Circular correlation with the real GDP." CESifo working paper number 3179, available online at: https://www.cesifo-group.de/de/ifoHome/publications/working-papers/CESifoWP/CESifoWPdetails?wp_id=14553255.

Anas J., Billio M., Ferrara L., Mazzi G.L. (2008), A system for dating and detecting cycles in the Euro area, The Manchester School, 76, 549-577.

Anas J., Ferrara L. (2004a). Turning points detection: the ABCD approach and two probabilistic indicators, Journal of Business Cycle Measurement and Analysis, No. 1, Vol. 2, pp. 1-36.

Anas, J. and Ferrara, L. (2004b): Detecting Cyclical Turning Points: the ABCD Approach and Two Probabilistic Indicators, paper presented at the 26th CIRET Conference, Taipei, October 2002, available online at: https://www.researchgate.net/publication/46557086_Detecting_Cyclical_Turning_Points_The_ABCD_Approach_and_Two_Probabilistic_Indicators.

Anas, J. and Mazzi G. L. (2015): A new graphical tool for business cycle monitoring, NTTS 2015 Conference Proceedings, European Commission, available online at: https://ec.europa.eu/eurostat/cros/system/files/NTTS2015%20proceedings.pdf.

Billio, M., Ferrara, L., Mazzi, G. L., Moauro, F. (2016): A multivariate system for turning point detection in the euro area, Eurostat Statistical Working Paper 2016, available online at: http://ec.europa.eu/eurostat/documents/3888793/7550218/KS-TC-16-007-EN-N.pdf/ec19307a-5dba-4e17-a081-d1ded4db632f.

Burns A. F., Mitchell W. C. (1946), Measuring business cycles. New York: NBER.

Leibfritz, W. und W. Nierhaus (1993), "Westdeutsche Wirtschaft: Wie tief ist die Rezession?", ifo Schnelldienst 46(7), 10-15.

Mazzi, G. L. (2015), Complementing scoreboards with composite indicators: the new business cycle clock, Eurona, 02-2015, available online at: https://ec.europa.eu/eurostat/cros/content/complementing-scoreboards-composite-indicators-new-business-cycle-clock-gian-luigi-mazzi_en.

Online Resources

Brazil (IBRE - Getulio Vargas Foundation and George Washington University): http://www.gwu.edu/~ibi/FGV%20Report%20Files/June2009.pdf

China (Business Cycle Signal System document): http://epp.eurostat.ec.europa.eu/portal/pls/portal/!PORTAL.wwpob_page.show?_docname=2304358.PDF

Eurostat: http://epp.eurostat.ec.europa.eu/cache/BCC2/group1/xdis_fr.html

Germany (Destatis): https://www.destatis.de/KoMo/Konjunkturmonitor.svg

Korea: http://kosis.kr/bcc/

Netherlands (CBS Statistics Netherlands): http://www.cbs.nl/en-GB/menu/themas/dossiers/conjunctuur/publicaties/conjunctuurbericht/inhoud/conjunctuurklok/conjunctuurklok2.htm

Malaysia (presentation document): http://www.bnm.gov.my/documents/conference_vol/2013_MyStats/np/Paper_Session%201_Abd%20Latib%20Talib%20%28DOSM%29.pdf

OECD: http://stats.oecd.org/mei/bcc/default.html

21

Comparing Alternative Composite Indicators for Euro-Area GDP

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Rosa Ruggeri Cannata
European Commission, Eurostat
rosa.ruggeri-cannata@ec.europa.eu

Table of Contents

21.1 Introduction

Gross domestic product (GDP) has for a long time been considered as the main macroeconomic variable giving an overall picture of the state of the economy. The growth of the production of goods and services reflects the expansions or contractions of the economy, and this kind of information is of high interest for policy makers and analysts. The relevance of the GDP as reference variable for the construction of composite indicators has been highlighted at several occasions in this handbook (see e.g. Mazzi and Calès, chapter 7 and Chapter 12, Gyomai et al. chapter 10 in this handbook). It is then particularly relevant to get timely and reliable estimates of GDP.

Unfortunately GDP estimates are the final step of a quite complex production process and are often available at quarterly frequency only, and with a significant delay. This is the reason why, analysts try to anticipate or estimate in real time main macroeconomic movements using advanced statistical and econometric techniques; such techniques can integrate official statistics mainly in two ways: rapid estimates or composite indicators anticipating GDP growth in real-time.

Since rapid estimates techniques, meaning the now/forecasting techniques applied to the recent past or the present, are out of the scope of this handbook, in this chapter we describe first several composite indicators aiming to anticipate GDP short run movements, and we propose then a real-time comparison of several indicators with respect to the Eurostat GDP flash estimate released around forty-five days after the end of the reference quarter. This comparison allows us to evaluate the relative performance of the indicators over the last years, which is an exercise of great interest since they have been characterised by a high degree of economic uncertainty. The list of composite indicators considered in this chapter is not exhaustive, but it is restricted to freely available indicators for the euro area plus some indicators internally developed at Eurostat.

Following this introduction, section 21.2 will introduce some general considerations about timeliness' improvement, section 21.3 will give an overview of freely available euro area leading indicators, while in section 21.4 we present the Eurostat GDP growth composite indicators, then section 21.5 presents the real-time simulation over the period from the third quarter of 2014 to the second quarter of 2016, finally section 21.6 concludes.

21.2 Improving GDP timeliness

When looking at how to improve the timeliness of GDP estimates, several feasible alternatives can be considered. Clearly the best solution would be to produce official estimates of GDP, but this would be feasible only by speeding up the production process. Usually, this process can be a quite long one and its improvement would require restructuring it and/or improve data availability at source level. The complexity is even greater to produce aggregates at euro area level because when full coverage is not reachable, the availability of estimates for member countries should guarantee that the aggregate estimate produced is robust enough to be considered as an official one. Often, thresholds in terms of coverage and revisions are first defined and then candidate estimates are tested for their performance, possibly on real-time data. Recently Eurostat has released a flash estimate of GDP for the euro area quarterly GDP thirty days after the end of the quarter; this chapter focuses on data available in a period when such rapid estimates were not yet produced.

Alternatively, fully model-based estimates can be considered; modelling would permit to produce real-time and advanced estimates. Since 2006 Eurostat has assessed several models to anticipate or estimate in real-time GDP movements either based on rapid estimates or on composite indicators techniques. An overview of these models can be found in Mazzi et al. (2016) and Mazzi, Moauro and Ruggeri Cannata (2016). However, models can perform in a different way under particular economic circumstances and phases, and the relative importance of the same variables in a given model can also change according to economic phases (see Mazzi, Mitchell and Montana (2014)). Based on the results of this wide comparative exercise, Eurostat has then decided to keep four alternative models leading to the construction of GDP growth composite indicators. Such

composite indicators have been constantly compared in real-time on a monthly base; they will be presented in section four.

21.3 Some freely available euro area leading indicators

Eurostat has undertaken a regular follow-up of a few euro area GDP growth indicators freely available. Several institutions do produce GDP growth estimates for the euro area but not all of them are freely available and accompanied by a clear methodological documentation. Actually Eurostat is monitoring two main indicators of GDP growth which are also regularly published in the Eurostatistics monthly bulletin: The Euroframe and the €-coin indicators. The indicators analysed by Eurostat mainly differ in terms of:

1. Release time;

2. Size and characteristics of the set of variables used;

3. Estimation procedure;

4. Coincident/leading behaviour.

In the following subsections we are providing a short description of the Euroframe and of the €-coin indicators.

21.3.1 The Euroframe Euro Growth Indicator

The **Euroframe Euro Growth Indicator** is compiled by the OFCE (Paris) in cooperation with the EUROFRAME group, which consists of the following institutions: CPB (Den Haag), DIW (Berlin), ESRI (Dublin), ETLA (Helsinki), IFW (Kiel), NIESR (London), PROMETEIA (Bologna), WIFO (Vienna), and CASE (Poland). The purpose of this leading indicator is to anticipate the development of the GDP in the euro area two quarters ahead of official statistics.

The indicator considers surveys from industry, construction, and households, and the real euro exchange rate. The indicator is based on a regression model in the log-differences of GDP with the following explanatory variables: production trend in industry, first difference, Construction confidence indicator, first difference and with a lag of five periods, Household's major purchases for the next twelve months, first difference and with a lag of one period, real euro/dollar exchange rate quarterly growth rate, two period lagged. The model includes two dummies for the fourth quarter 2008 and the first quarter 2009. The parameters of the regression model are estimated by ordinary least squares (OLS). The estimation period starts in 1993 and the regression parameters are re-estimated each time. GDP estimates of the two quarters ahead are obtained as a two-step forecast of the regression model. GDP growth rate estimates are released on a monthly basis at the beginning of each month.

21.3.2 The €-coin indicator

€-coin is a real-time, monthly estimate of euro area-wide GDP growth, computed each month by the staff of the Banca d'Italia. It provides a single number summarizing the current economic picture for the euro area. €-coin collates a large collection of statistical data covering around 150 series and referring both to the euro area as well as to its major economies, from June 1987. The input dataset includes groups of variables that are leading, lagging and coincident with respect to the GDP. The series stem from industrial production, business surveys, stock market and financial data, demand indicators, prices, money aggregates, interest rates, financial variables, trade variables, labour market and daily series (stock markets).

€-coin is obtained as a projection of the (band-passed) GDP on factors derived from a generalization of the concept of principal components. Variables belonging to the input dataset that are leading with respect to the GDP act as proxies for future values of GDP missing at the end of the sample. The method is based on the computation of linear combinations, a generalization of ordinary principal components, referred to as "smooth factors", where the linear combination is chosen in a way to reach an optimal compromise between smoothness and goodness of fit. For more details about the methodology see references listed in the bibliography.

Two technical issues had to be dealt with in the construction of the indicator; the first concerns the availability of series at quarterly frequency only, when monthly data were needed; techniques to obtain monthly observations range from linear interpolation to more sophisticated temporal disaggregation techniques. However, being the focus on a series obtained by removing short-run oscillations, the particular method employed to interpolate values within quarters did not have a relevant impact, so that linear interpolation has been retained. The second technical issue concerns the different delays in data availability; in this case, several methods are considered and monitored varying from simply shifting forward the available values to predict missing values using ARMA models or the EM algorithm. **€-coin** is published at the end of each month and refers to the same month; the release is accompanied by an assessment on how the release of new data affects the real-time estimates of underlying GDP growth in the euro area, comparing the final official estimate and the preliminary estimate, an one step ahead forecast.

21.4 Eurostat coincident indicators of GDP growth

Since 2006 Eurostat, within the activities related to the improvement of Principal European Economic indicators (PEEIs), started to compile three competing models for the GDP nowcasts. Two out of the three models are based on factor models, while the third one is based on a bridge regression model. These models have been extensively described in Charpin, Matthieu and Mazzi (2016) and Charpin (chapter 18), and in Mazzi et al. (2016). The models were run monthly to produce estimates for the GDP growth of the previous quarter; for each quarter three estimates of GDP were produced: the first at the end of the second month (T-30), the second at the end of the quarter (T+0) and the last one a month later the end of the quarter (T+30).

The selected models are based on regressions techniques using either individual series or principal components as regressors. Principal components (PC) are usually extracted from a large data set of coincident and leading series, all entering the data set without any lag; then the most important PCs are introduced in a regression model, possibly with lags. Considering that the introduction of many series, more or less related to GDP, could produce a noise that deteriorates the estimate (see Boivin and Ng (2006)), the model selects only series directly related to GDP growth, in principle series that can help to predict GDP growth but could not be introduced simultaneously in a regression because of multicollinearity. Moreover, these series are lagged if they show leading properties in regression models. Thus, principal component regressions can be viewed here as a way to solve the multicollinearity problem.

The selected regressors (individual series or principal components) can be classified into two groups, i.e. coincident or leading variables. Leading variables enter the regression with at least one lag and are thus entirely available at the date of the estimation. The inclusion of coincident regressors raises a difficulty because they are not always available when producing the estimate; hence they have to be forecasted. Coincident regressors will be chosen among survey data for their timely availability, with the exception of industrial production. Industrial production is a good candidate among explanatory variables being a good proxy of gross value added in the industry, which is a relevant component of GDP and is used by many euro area countries to produce flash estimates. When producing a coincident GDP indicator for quarter T, the missing months for industrial production in the quarter (3, 2 or 1) are forecasted by a regression model. Concerning survey data, at most one month is missing for the first GDP estimate, in this case the average of the two available months is used as an estimate of the three month average.

Concerning data selection, the approach is based on a modified version of the LARS algorithm introduced by

Bai and Ng (2008). The LARS algorithm choses targeted predictors (i.e. the most appropriate variables to estimate the quarterly GDP) in a less selective way with respect to other stepwise algorithms; in fact, it allows retaining correlated series which is a desirable property in view of extracting principal components. Time series are not discarded, they are simply ranked by decreasing predictive power according to the selection criterion of the LARS algorithm.

The modification applied concerned several aspects. Firstly, some series were introduced several times in the data set with different lags, in order to better exploit the coincident or leading features of the series in the LARS algorithm. Secondly, soft data were considered both in levels and variation to improve their forecast capabilities. The LARS algorithm is able to rank rapidly a large number of series with respect to their predictive ability. Moreover, the LARS algorithm was applied to a dataset not containing financial series; after the extraction of the principal components, financial series were introduced directly in the regression, based on the consideration that often financial series generate additional principal components.

The three models differ either by the kind of series involved or by the statistical methodology used.

Later on Eurostat has also developed a fourth model, which is a disaggregated model for nowcasting GDP based on a VAR specification for a number of Member States, using the results as an input for a euro area estimate. For technical detail about this model please refer to: Charpin, Mazzi and Moauro (2016). In the following subsections we briefly recall the main features of the models mentioned above.

21.4.1 The factor model based on soft information (FS model)

This factor model is based on factors constructed exclusively with soft data; it has been built in order to estimate coincident GDP growth when no hard data is available for the current quarter which is the case of the estimation carried out at T-30 days from the end of the reference quarter. This model includes thirteen survey data plus the euro/dollar real exchange rate in growth rates. Surveys data cover the Industry, Consumers, Construction, and Retail trade sectors.

21.4.2 The factor model base on hard and soft information (FHS model)

This factor model has been built with both soft and hard data (FHS model). It includes fourteen series of which nine surveys data covering the Industry, Consumers, Construction, and Retail trade sectors plus five hard data: Industrial Production (excluding construction), Construction production, Exports, Retail sales and Unemployment rate.

21.4.3 The bridge model based on hard and soft information (BHS model)

The third considered model is a bridge model containing hard and soft data (BHS model). The bridge model includes nine variables of which five hard: Industrial production, Construction output, Exports, Retail sales, Unemployment rate, and three survey data: Consumer opinion over the next twelve months, Employment expectations in construction, and Construction confidence indicator, plus the euro/dollar real exchange rate.

21.4.4 The disaggregated model

Estimates of a geographical aggregate can also be indirectly obtained by aggregating component estimates, this is called an indirect approach. In this subsection we are shortly illustrating the indirect approach to estimates GDP growth. Vector autoregressions (VAR), with cointegration terms when needed, have been

estimated for the following countries: Germany, Italy, France, Belgium, the Netherlands and Spain. The models have been based on the accounting identity defining GDP as the sum of Private consumption, Public consumption, Investment, Inventory changes and Exports minus Imports, which is showing a potential cointegration relation among GDP and some demand components available at monthly frequency; when this was not the case, proxy variables where used. Since the monthly information used in bridge models is published ahead of GDP, it can be considered as exogenous and an individual error correction equation is sufficient for the model. Thus we can estimate an individual error correction equation, instead of a system of equations.

If y denotes the logarithm of GDP, x_h the logarithm of the h hard data (quarterly converted) and z_k some $I(0)$ variable, the error correction model in equation 21.1 is estimated by:

$$\Delta y = c + \alpha y_{-1} + \Sigma_h \alpha_h x_{h,-1} + \Sigma_h b_h \Delta x_h + \Sigma_k d_k z_k \qquad (21.1)$$

The cointegration test is carried out with the T-ratio of coefficient α and the Ericsson-Mackinnon Table (2002).

When cointegration is not accepted, the usual bridge model prevails, as described in equation 21.2:

$$\Delta y = c + \Sigma_h b_h \Delta x_h + \Sigma_k d_k z_k \qquad (21.2)$$

The missing observations of hard data in quarter T are forecasted using auxiliary monthly regressions (which contain autoregressive terms and survey data).

For the euro area and its main countries, the potential variables entering the cointegration relation are: Retail sales, Consumption in manufactured goods (in France), Exports and Imports (with a negative sign). Other $I(0)$ variables, z_k, can be the Industrial Production Index (IPI) growth rate, the Construction output growth rate, the Change in unemployment rate, survey data, the euro/dollar real exchange rate in growth rate, and financial data (interest rates, stock index). Some variables are lagged especially exchange rates, financial data and some survey data.

The main characteristics of individual country models and the euro area one are presented below. No cointegration relation has been found for Italy, Spain and Belgium where usual bridge models are estimated. In almost all our country models foreign trade plays an important role, most frequently through exports but for Germany through imports also.

France

The cointegration relation for France embeds the logarithms of GDP, Consumption in manufactured goods, and Exports. Other variables of the right side of the model are:

  (i) the log-difference of Industrial production, of Consumption in manufactured goods, of Exports, and of Real MSCI stock index;

 (ii) the first difference in Opinions on the ability to save (two lags, French consumer survey) and on the Probable output trend (French business survey); and the first difference of the Construction confidence indicator (two lags, EU construction survey).

The cointegration is strong (the T-ratio is equal to -6.8, the 5% critical value being -3.5) on the pseudo real-time evaluation period 2007-2011, and the root mean square error is equal to 0.19 percentage points. High cointegration can be explained by the availability of monthly Consumption in manufactured goods, a substantial part of households' consumption, which is known as a driving factor of growth in France.

Germany

The cointegration relation for Germany embeds GDP, Exports and Imports, all expressed in logarithms. In this equilibrium relation, Exports have a positive coefficient whereas Imports have a negative coefficient. In Germany, Exports are the main driving factor of growth. Other variables of the right side of the model are:

(i) the log-difference of Industrial production, of Construction output, of Exports and of Retail sales;

(ii) the first difference in Opinions in consumer survey concerning the expected saving situation over the next twelve months and Major purchases over the next twelve months (one lag); and the Short-term interest rate in first difference (one lag).

Co-integration is accepted only at a significance level of 10%. Nevertheless, this model has been retained because gives unbiased estimates and a better accuracy of other models used for comparison. The model has been assessed over the period 2007-2011. Although the root mean square error is equal to 0.34 percentage points, this reflects the higher GDP volatility of GDP growth.

The Netherlands

The cointegration relation embeds the logarithms of GDP, Retail sales and Exports. Other variables of the right side of the model are:

(i) the log-difference of Industrial production, of Retail sales, and of Exports;

(ii) the first difference in Opinions on order book position in construction, on the Activity compared to last months (construction survey), and Households' opinion on the economic situation.

Co-integration is almost accepted at a significance level of 5%. The model has been assessed over the period 2007-2011, with a root mean square error equal to 0.34 percentage points, the model accuracy could then be improved.

Italy

No cointegration relation has been found for Italy. The Italian GDP quarterly growth rate is then obtained using a bridge model. The GDP growth rate is determined by the Industrial production index (growth rate), Real exports (growth rate), euro/dollar real effective exchange rate (growth rate, three lags), Opinion of industrials on production expectation for months ahead (level), Opinion of industrials on stocks (change), Opinion of retailers on stocks (level, one lag), and Households' opinion on major purchases at present (change, one lag). The standard error of estimates is equal to 0.24 percentage points whereas the standard error of the GDP growth rate is equal to 0.67 percentage points over the period 2007-2011.

Spain

No cointegration relation has been found for Spain. Thus the Spanish GDP quarterly growth rate is obtained using a bridge model. Among available hard data, only the change in the Unemployment rate is significant in the GDP growth equation. The other variables entering the equation are the Opinion on employment (Retail trade survey, one lag), the change in Opinion on export order book position (Industrial survey, one lag), the change in Opinion on the financial situation over next twelve months (Consumer survey, one lag), the euro/dollar real exchange rate (growth rate, one lag), and the real Oil price (growth rate, five lags). The standard error of the estimate is 0.14 percentage points whereas the standard error of the GDP growth rate is 0.60 percentage points over the period 2007-2011.

Belgium

No co-integration relation has been found for Belgium. Thus the Belgian GDP quarterly growth rate is obtained using a bridge model. Among available hard data, only Exports is significant in the GDP growth equation. Estimates for Belgium concern T-30 and T+0 only because Belgium releases GDP growth values at T+30.

Euro area direct model

In the euro area model, the cointegration relation embeds GDP, Retail sales and Exports logarithms. Other variables of the right hand side of the model are the log-differences of Industrial production, Construction output, Exports, Unemployment rate, euro/dollar real exchange rate (two lags), and the difference of Opinions concerning the present business situation (Retail trade survey), the Major purchases over the next twelve months (one lag, consumer survey). Over the period 2007-2011, the root mean square error is equal to 0.22 percentage points and cointegration is only significant since the fourth quarter of 2009, at a 10% level.

Euro area indirect model

We illustrate now how the euro area GDP growth is obtained by using country available results. We derive the euro area GDP estimates by using the six GDP estimates within a regression model, which returns the weights of each country GDP on the total euro area GDP growth. The regression has a constant term and is estimated by the Cochrane-Orcutt method if residuals have an autocorrelation of order one. Otherwise, an OLS estimator is applied. The estimation period starts in the second quarter of 1995. This regression is run at each estimation date. Comparing this model with the euro area direct one, it was found that GDP fluctuations are well and much better tracked by the indirect model. We will then discard the direct model in the comparative analysis presented in the rest of the chapter.

21.5 Real-time assessment

In this section we compare the composite indicators introduced in the previous two sections. In order to make the simulation exercise clearer and better understandable by readers the following elements should be taken into account:

- The reference month for the Euroframe and the €-coin indicators is attributed on the basis of the publication in the Eurostatistics bulletin, which not necessarily corresponds to their releases.

- Couple of values for the €-coin indicator appear missing since they were not available at the moment of the Eurostatistics bulletin release.

- The first three euro area models proposed by Eurostat in the previous section have been estimated by using real-time vintages so that the exercise is fully real-time.

- The indirect model has been estimated only in a pseudo real-time way, due to unavailability of proper vintages for few national variables.

Recently Eurostat started to publish official GDP estimates at T+30; the T+30 estimates from the various models are then losing their relevance. Nevertheless we have kept them in this chapter for a number of reasons. The first is that the Eurostat official GDP estimates were still released at T+45 during most of the period covered by the comparative exercise. The second one is that it is also of great interest to evaluate the performance of model-based estimates against an official one, even in the case they are released at the same time. Finally, it is important to notice that, since between T+15 and T+30 days there are no significant data releases affecting the models presented in section four, they could also be anticipated and computed at T+15

days after the end of the reference quarter, regaining then in relevance in comparison with the official T+30 GDP estimate.

In table 21.1 below we compare the estimates of the Euroframe and €-coin indicators, introduced in section three, as available in different months of the current and next quarter, with the official flash estimate produced by Eurostat at T+45. Looking at those three releases, we are able to compare them with those produced by the Eurostat models, coupling the different months to the T-30, T+0 and T+30 days after the end of the quarter.

Table 21.1: Real-time values Euroframe and €-coin quarterly growth

	Quarter T	Euroframe q/q-1	€-coin q/q-1	Eurostat flash estimate T+45
August 2014	2014 Q3	0,40	0,30	
September 2014		0,20	0,20	0,16%
October 2014		0,20	0,20	
November 2014	2014 Q4	0,20	0,10	
December 2014		0,30	0,10	0,34%
January 2014		0,30	0,10	
February 2015	2015 Q1	0,40	0,20	
March 2015		0,30	0,20	0,40%
April 2015		0,40	0,20	
May 2015	2015 Q2	0,60		
June 2015		0,60	0,40	0,31%
July 2015		0,60	0,40	
August 2015	2015 Q3	0,70		
September 2015		0,60	0,40	0,30%
October 2015		0,50		
November 2015	2015 Q4	0,70	0,40	
December 2015		0,70	0,40	0,27%
January 2016		0,70	0,40	
February 2016	2016 Q1	0,50	0,50	
March 2016		0,70	0,50	0,52%
April 2016		0,30	0,40	
May 2016	2016 Q2	0,30	0,30	
June 2016		0,50	0,30	0,25%
July 2016		0,50	0,30	

Source: Authors' own calculations

In table 21.2, we compare the estimates generated by the various Eurostat models at T-30, T+0 and T+30 introduced in section four, with the Eurostat T+45 flash estimate. The comparisons are performed over the period from the third quarter of 2014 to the second one of 2016. Being the assessment period quite short, the following analysis has to be considered as indicative and should be extended to a larger number of observations before arriving to more conclusive results.

Table 21.2: GDP growth estimates according to the different models

End of quarter T		Estimates			Eurostat flash estimate
		T-30	T+0	T+30	T+45
2014 Q3	Factor soft	0,40%	0,39%	0,39%	
	Factor Hard & Soft	0,29%	0,19%	0,17%	
	Bridge	0,34%	0,20%	0,10%	0.16%
	Indirect	0,28%	0,31%	0,22%	
2014 Q4	Factor soft	0,42%	0,42%	0,43%	
	Factor Hard & Soft	0,28%	0,35%	0,46%	
	Bridge	0,23%	0,30%	0,38%	0.34%
	Indirect	0,23%	0,29%	0,30%	
2015 Q1	Factor soft	0,30%	0,35%	0,35%	
	Factor Hard & Soft	0,42%	0,50%	0,54%	
	Bridge	0,41%	0,47%	0,46%	0.40%
	Indirect	0,32%	0,51%	0,45%	
2015 Q2	Factor soft	0,40%	0,39%	0,39%	
	Factor Hard & Soft	0,46%	0,45%	0,35%	
	Bridge	0,35%	0,46%	0,40%	0.31%
	Indirect	0,38%	0,54%	0,47%	
2015 Q3	Factor soft	0,59%	0,61%	0,61%	
	Factor Hard & Soft	0,46%	0,53%	0,49%	
	Bridge	0,39%	0,51%	0,41%	0.30%
	Indirect	0,54%	0,57%	0,43%	
2015 Q4	Factor soft	0,67%	0,73%	0,66%	
	Factor Hard & Soft	0,47%	0,37%	0,43%	
	Bridge	0,42%	0,38%	0,46%	0.27%
	Indirect	0,46%	0,40%	0,41%	
2016 Q1	Factor soft	0,48%	0,45%	0,45%	
	Factor Hard & Soft	0,43%	0,66%	0,55%	
	Bridge	0,36%	0,67%	0,45%	0.52%
	Indirect	0,57%	0,70%	0,39%	
2016 Q2	Factor soft	0,48%	0,49%	0,49%	
	Factor Hard & Soft	0,39%	0,29%	0,36%	
	Bridge	0,40%	0,31%	0,32%	0.25%
	Indirect	0,20%	0,21%	0,17%	

Source: Authors' own calculations

In this analysis, we look at the concordance sign, the mean forecasting error (MFE) and the root mean squared forecasting error (RMSFE). The first finding is that all models show at each estimation point a perfect sign concordance with the Eurostat official release. Looking at the mean forecasting error (MFE) with respect to the Eurostat official flash estimate over the evaluation period, we observe that the Eurostat bridge and the €-coin models perform in a very similar way; €-coin is the model with the minimum bias since its MFE is very low; moreover, €-coin is the only indicator with a slight tendency to underestimate the GDP growth. However, the Eurostat indirect model outperforms €-coin at T+30, and the bridge model's mean forecasting error is quite close too, so that we could say the three models are equivalent at T+30. Almost all the models show a deterioration of the performance at T+0 with respect to T-30, followed by a slight improvement at T+30. This is probably due to the high volatility of some hard series, such as the industrial production, which tend to lower

the forecasting ability when entering for the first time in the model.

When looking at the RMSFE with respect to Eurostat official flash estimate over the considered evaluation period, Eurostat bridge model and the factor model with hard and soft data outperform €-coin at the three release dates; moreover, all Eurostat models perform better than the Euoframe one in the evaluation period in terms of RMSFE. Concerning the indirect model, it shows a less satisfactory performance at T+0 while at T-30 and T+30 it is almost in line with other well performing models. When considering Euroframe and the Eurostat factor with soft data only, it appears that they perform in a similar way but they are clearly outperformed by the other models.

When considering how to improve the models, we could focus on the indirect one; enhancements of the country specific models and/or enlarging the countries' coverage could potentially benefit to the indirect model. Finally, we would like to stress that all the models introduced in this chapter perform quite well at T-30, showing that a relevant improvement in timeliness thanks to econometric techniques is feasible.

21.6 Conclusions

In this chapter we have introduced several model-based estimates of GDP growth aiming to improve timeliness; the use of modelling techniques to complement official statistics is of relevance when considering users' needs, in particular for policy makers. The models introduced refer to either publically available composite indicators such as Euroframe (OFCE, France) and €-coin (Bank of Italy), or from an Eurostat set of composite indicators developed for internal purposes only.

A large variety of models has been considered, covering static and dynamic factor models, regression models taking in consideration mixed frequencies in the input data such as the bridge one, as well as VAR and VECM models, for more detail about those models see Luciani, chapter 17 and Charpin, chapter 18 of this handbook.

Various estimates are calculated at different points in time reflecting the data flow and availability, with surveys data being already available at the end of the quarter while hard data are generally released later. The analysis has shown that results are quite satisfactory with a slight deterioration for the estimates carried out right at the end of the quarter. Despite the short simulation period, the results show that estimates carried out one month before the end of the quarter appear of quite good quality, taking into account their lead with respect to official estimates. Finally, the already very good performance of the indirect model described in section four shows that increasing the country coverage for this approach and enhancing country specific models could likely lead to even better results.

Bibliography

€-coin online tutorial available online at: http://eurocoin.cepr.org/files/eurocoin_tutorial.pdf

Altissimo, F., Cristadoro, R., Forni, M., Lippi, M. and G. Veronese (2008), New Eurocoin: Tracking Economic Growth in Real Time, The Review of Economics and Statistics, No. 92, 1024-1034, available in the Bank of Italy working paper series (Temi di Discussione della Banca d'Italia) no. 631, available online at: http://www.bancaditalia.it/pubblicazioni/temi-discussione/2007/2007-0631/en_tema_631.pdf

Bai J., Ng S. (2008), Forecasting Economic Time Series Using Targeted Predictors, Journal of Econometrics 146, 304-317.

Boivin J., Ng S. (2006), Are more data always better for factor analysis?, Journal of Econometrics, vol 132, issue 1, May, 169-194.

Charpin F. (2017), Chapter 18: Regression, principal components and small scale factors based models, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Charpin F., Mathieu C., Mazzi G.L. (2016), Predicting volatility: Construction of coincident indicators for the euro area, Eurostat Statistical Working papers, KS-TC-16-012-EN-N.

Charpin F., Mazzi G.L., Moauro F. (2016), Early estimates of GDP growth rates by error correction models, Eurostat Statistical Working papers, KS-TC-16-011-EN-N.

Gyomai G., Ahmad N., Astolfi R. (2017), Chapter 10: The OECD System of Composite Leading indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Luciani M. (2017), Chapter 17: Large-Dimensional Dynamic Factor Models in Real-Time, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mazzi G.L., Calès L. (2017), Chapter 7: An overview of cyclical indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mazzi G.L., Calès L. (2017), Chapter 12: An overview of alternative turning points composite indicators, in: Handbook on Cyclical Composite Indicators, ed. by G.L. Mazzi.

Mazzi G.L., Moauro F., Montana G., Ruggeri Cannata R. (2016), Towards a real-time system for the analysis of the euro area cyclical situation, Eurostat Statistical Working papers, KS-TC-16-004-EN-N.

Mazzi G.L., Mitchell J., Montana G. (2013), Density nowcasts and model combination : nowcasting Euro-area GDP growth over the 2008-9 recession, Oxford Bulletin of Economics and Statistics, Volume 76 (Number 2), 233-256,

Mazzi G.L., Moauro F., Ruggeri Cannata R. (2016), Advances in econometric tools to complement official statistics in the field of Principal European Economic Indicators, KS-TC-16-013-EN-N.

VII | Guidelines

22 Guidelines for the Construction of Composite Cyclical Indicators

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Ataman Ozyildirim
The Conference Board, Inc.
ataman.ozyildirim@conference-board.org

Table of Contents

22.1 Introduction

22.1.1 Motivation for the guidelines

These guidelines have been developed as the ideal conclusion of the handbook to promote the use best practices in the construction of cyclical composite indicators to:

- Support the construction of a harmonised system of cyclical composite indicators at national and regional level

- Enhance comparability of cyclical composite indicators across countries and regions

- Increase the effectiveness of business cycle analysis.

22.1.2 Scope of the guidelines

These guidelines are firstly addressed at statistical institutions already involved in the compilation of cyclical composite indicators or those that are evaluating, establishing their programs in this field. The topics covered in the guidelines and the recommendations proposed make them also of high interest for public and private institutions working on the compilation of cyclical composite indicators.

Furthermore, the guidelines are aimed at anyone whose work is related in some way to the compilation of cyclical composite indicators. The guidelines have been conceived both for experts and beginners.

The guidelines cover all issues related to the construction, the regular production and the maintenance of cyclical composite indicators. They cover a range of issues from data problems through compilation aspects of various kinds of composite indicators to revisions. The guidelines are restricted to macro-economic cyclical composite indicators so that they do not cover sectoral cyclical indicators. However, some of the recommendations and suggestions provided here can be relatively easily adapted to the case of sectoral composite indicators.

The guidelines are based on a set or principles, presented in the annex to this introduction, which give some general rules to be followed when compiling cyclical composite indicators.

The guidelines are structured in seven sections, each of them dealing with a specific issue or indicators typology such as, data problems, compilation of turning points composite indicators etc. Each section is subdivided in a number of items dealing with a specific aspect. All items are presented following a standard template subdivided into three parts: a description, a list of options and a list of ranked alternatives.

The options list, without pretending to be exhaustive, the various possibilities to deal with the specific problem treated in the item. Out of these options, three ranked alternatives are highlighted:

(A) Best alternative

(B) Acceptable alternative

(C) Alternative to be avoided

The best alternative (A) should always be the feasible target for producers. It should always be achievable with a reasonable effort, unless some production or institutional constraints prevents it.

The acceptable alternative (B) should be viewed as an intermediate step towards the achievement of alternative A. It could also be seen as the target for a limited number of cases when specific data issues, user requests, time or resource constraints prevent the achievement of the alternative (A).

The alternative to be avoided (C) includes our recommendations against some procedures.

The objective of the guidelines is helping producers of cyclical composite indicators to move towards alternative (A). Careful considerations and, possibly, prompt measures should be taken whenever alternative C is in use.

22.1.3 Advantages and drawbacks

Implementing recommended guidelines will enhance the transparency and clarity of a system of cyclical composite indicators. It will also increase the cross-country/region comparability and will contribute to the development of an effective real-time monitoring system of the short-term economic conditions. On the other hand, the adoption of the guidelines will require additional efforts by the institutions in charge of compiling cyclical composite indicators to progressively comply with the requirements. The decision of not applying the guidelines will prevent at least partially the process of producing more and more comparable cyclical composite indicators to facilitate a global view of the short-term economic conditions.

22.1.4 Main principles for the compilation of cyclical composite indicators

○ Principle 1: Objectivity and impartiality: Cyclical composite indicators should deliver messages based on statistical evidence coming from data without any a priori conditioning of any economic or political theory.

○ Principle 2: Methodological soundness: Cyclical composite indicators should be complied according to sound, well tested, international agreed and recognised methods.

○ Principle 3: Clarity: Cyclical composite indicators should be accompanied by a clear and complete documentation easily accessible and understandable also by non-expert users.

○ Principle 4: Transparency: All steps of the compilation process of cyclical composite indicators should be clearly listed, described and documented.

○ Principle 5: Interpretability and readability: The cyclical composite indicators should be easily interpretable and readable even by non-expert users. This can be achieved also by complementing cyclical compost indicators with visualisation and graphical tools to facilitate the delivery of the messages.

○ Principle 6: Consistency: Various kinds of cyclical composite indicators compiled by the same institution should deliver consistent messages. Possible inconsistency, when it happens, should be highlighted and, explained and, whenever possible, corrective measures should be undertaken without violating the other principles.

○ Principle 7: Comparability: All institutions involved in the compilation of cyclical composite indicators should foster the comparability of their indicators at national level, across countries and regions.

22.2 General policy

22.2.1 Statistical institutions and their role in compiling cyclical composite indicators

Description

The compilation of cyclical composite indicators has been always considered borderline tangential subject for statistical institutions. The attitude of statistical institutions vis-à-vis of this problem is quite articulated and continuously evolving. Some European statistical institutes such as INSEE and ISTAT are traditionally quite open on this subject as it also the case for Statistics Netherlands. Others remain much more sceptical. Outside Europe, the situation is very articulated: while some emerging and developing countries are very active in the field of cyclical composite indicators, other statistical institutes such as those in North America ones remain opposed. These attitudes depend on their perspectives on the clear separation between data producers and data users and on the separation between data and their interpretation.

If cyclical composite indicators can be considered as part of the statistical production, even if not necessarily part of official statistics, then statistical institutes can play an active role in this field; otherwise they should stay away. The involvement of statistical institutes in the compilation of cyclical composite indicators will have the big advantage that official statisticians are the ones who know better data characteristics and specificities and this knowledge can be used for improving the quality of cyclical composite indicators. On the other hand, statistical institutes can contribute to enhancing the transparency and objectivity of cyclical composite indicators.

Options

- Direct involvement of statistical institutes in compiling and disseminating cyclical composite indicators

- Cooperation of statistical institutes with other institutions involved in the compilation and dissemination of cyclical composite indicators

- Statistical institutes not involved in the compilation of cyclical composite indicators

Alternatives

– Alternative A: Conduct an in-depth investigation of the advantages and drawbacks, costs and benefits of the direct involvement of statistical institutes in the production and dissemination of cyclical composite indicators and make public all available documents explaining the results on the position of the statistical authority;

– Alternative B: Analyse advantages and drawbacks, costs and benefits or cooperating with other institutions in charge of the production and dissemination of cyclical composite indicators and make the results publicly available;

– Alternative C: No involvement in the compilation of cyclical composite indicators.

22.2.2 A general policy for the compilation of cyclical composite indicators

Description

A general policy for the compilation of cyclical composite indicators should define the main compilation strategy, based on the set of principles presented in the annex. At a minimum, it should contain information related to: the type of cyclical composite indicators to be compiled, the adopted methodological and quality framework, and the revision and dissemination strategy for such indicators.

Options

- A clear and comprehensive policy based on a set of agreed principles

- A policy only partially based on a set of agreed principles

- A policy not in line with a set of agreed principles

- No general policy

Alternatives

- Alternative A: Elaborate and disseminate a general policy for the compilation of cyclical composite indicators, based on a set of agreed principles, and detailing all key aspects related to the production, revision, quality assessment and dissemination of such indicators;

- Alternative B: Elaborate and disseminate a policy for the compilation of cyclical composite indicators, only partially in line with a set of agreed principles, covering at least partially all key production, revision, quality assessment and dissemination aspects for such indicators;

- Alternative C: Incomplete or not in line with agreed principles policy or lack of any policy.

22.2.3 Stability of the general policy

Description

Keeping the general policy for cyclical composite indicators stable over the time will increase users' confidence and will allow producers to work within a stable framework. On the other hand, assuming that the general policy should never be revised is not realistic since the economic and statistical conditions will evolve and consequently policies should be updated.

Options

- keep the general policy for cyclical composite indicators stable over the time as much as possible and make only changes that are announced in advance after careful vetting;

- change the general policy for cyclical composite indicators with a fixed calendar (i.e. each 5 years);

- change the general policy often, like each year, to incorporate all new elements;

- change the general policy on an irregular basis.

Alternatives

- Alternative A: Keep the general policy for the compilation of cyclical composite indicators as much as possible stable over the year and adapt it only when significant changes occur either in the economic system or in the statistical system; changes to the general policy should be preannounced;

- Alternative B: Update the general policy on a fixed horizon allowing for a sufficient stability of the policy (for example every 5 years);

- Alternative C: Update the general policy too often or update it irregularly and without any pre-announcements.

22.2.4 Quality framework for cyclical composite indicators

Description

Quality assessment of cyclical composite indicators needs to consider all five dimensions of statistical output quality, as listed in the European Statistics Code of Practice:

- Relevance

- Accuracy and reliability

- Timeliness and punctuality

- Coherence and compatibility

- Accessibility and clarity

Measures identified for each dimension can be qualitative or quantitative — qualitative measures will normally be "Yes" or "No", and quantitative measures will normally be test statistics with a direct interpretation of "Pass" or "Fail".

For example, relevance is measurable qualitatively through consultation with users, while accuracy and reliability are measurable quantitatively.

Since cyclical composite indicators are somewhat different from official statistics, the quality framework needs to be adapted to incorporate their specificities without changing its main spirit.

Options

- Use the quality framework for official statistics duly amended to incorporate specific features of cyclical composite indicators;

- Use the default quality framework for official statistics;

- Use an ad hoc quality framework for cyclical composite indicators;

- Do not use any quality framework.

Alternatives

– Alternative A: Use a quality framework fully in line with the one in use for official statistics but customised to take into account specificities of cyclical composite indicators;

– Alternative B: Use the default version of the quality framework in use for official statistics;

– Alternative C: Use an ad hoc quality framework or do not use a quality framework.

22.3 Data issues

22.3.1 Data availability and quality

Description

The availability of a large set of macro-economic statistical indicators available at infra-annual frequency (possibly and preferably monthly) is a crucial pre-requisite for the compilation of cyclical composite indicators. These macroeconomic statistical indicators should meet some important requirements:

- Availability of a sufficiently long time period (covering at least two or three cycles, i.e. 15 to 18 years)

- Regular behaviour with a limited number of outliers, structural changes and other irregularity

- Compiled according to internationally agreed standards

- Regularly produced and disseminated on the basis of a clear release calendar

- Accompanied by a clear documentation (i.e. metadata)

The compilers of cyclical composite indicators should be able to regularly monitor the overall quality of the statistical indicators (i.e. component indicators) they are going to use in their activity of constructing composite indicators. This monitoring will allow them to have the best possible knowledge on the data, of their generating process as well as of their advantages and drawbacks.

Options

- Developing a small-sized database containing relevant macro-economic statistical indicators available at infra-annual frequencies, and implementing on it an automatic system of quality monitoring to be used periodically such as every week or month;

- Regularly following up the quality of macroeconomic statistical indicators using monitoring systems already in place, if any, or conducting a non-automatic and non-exhaustive assessment;

- Performing only irregularly an assessment of macro-economic statistical indicators;

- Do not perform any specific quality assessment.

Alternatives

- Alternative A: Develop and maintain a small database for macroeconomic statistical indicators potentially relevant for he compilation of cyclical composite indicators and implement on it an automatic system of quality monitoring and assessment to be run on a regular basis (ideally every month);

- Alternative B: Follow up the quality of relevant macro-economic statistical indicators by using all already available quality information and complement it with additional checks conducted on a regular basis but in a non-automatic way;

- Alternative C: Perform a non-regular quality control not necessarily covering all relevant statistical indicators or do not perform at all any quality control.

22.3.2 Real-time vintages

Descriptions

The construction of cyclical composite indicators could benefit a lot from the availability of real-time vintages of most relevant macro-economic statistical indicators. Historical vintages of as-published data will allow compilers of cyclical composite indicators to conduct comparative simulated tests in real-time to assess the performance of alternative composite indicators in order to select the best performing one. Developing and maintaining a system for the automatic collection and storage of vintages is not an easy task and it requires a certain investment on IT and human resources by statistical institutes which are, obviously, the best places to execute this task. Several national and international organisations, both public and private ones, have invested during the last two decades in this field and the amount of vintage databases has considerably grown. Nevertheless, the coverage across many countries remains still unsatisfactory.

Options

- Already collecting and storing in an appropriate IT framework vintages from most relevant macro-economic statistical indicators;

- Developing a new system for collecting and storing vintages on a regular basis;

- Collecting vintages on a purely voluntary basis without having an appropriate IT storage system;

- Using national vintages already collected by other institutions;

- Do not collect or plan to collect any vintage.

Alternatives

- Alternative A: Use an already existing system for data-vintages collection and storing on regular basis or developing it to ensure vintages availability for the future;

- Alternative B: Collecting vintages in a voluntary basis without a good infrastructure for the storage or using already existing vintages collected by other institutions;

- Alternative C: Do not collect or plan to collect data-vintages.

22.3.3 Unavailability of seasonally adjusted data

Description

The presence of seasonal fluctuation can hide, at least partially, the cyclical component and impair its proper identification and forecasting. This is the reason for which, when constructing cyclical composite indicators, there is usually a strong preference for using seasonally adjusted data for those statistical indicators which appear to be affected by season movements. Unfortunately, statistical institutes do not necessarily provide a full set of seasonally adjusted statistical indicators in all cases. In some cases, they do not even provide any seasonally adjusted data. This is especially true for developing economies and even some emerging market economies which still have a very basic statistical system. But it can happen also in more mature economies. Filling this gap is one of the main tasks to be accomplished before starting the construction of cyclical composite indicators.

Options

- All relevant statistical indicators, showing the significant presence of seasonal movements, are made officially available also in the seasonal adjusted form either by statistical institutes or by other national or international institutions a great deal of experience in the field, such as central banks, the IMF etc.;

- Only few statistical indicators are officially unavailable in the seasonally adjusted form;

- All statistical indicators are officially unavailable in the seasonal adjusted form;

- A number of statistical indicators is made available in the seasonal adjusted form by private institutions or academics without any official stamp.

Alternatives

- Alternative A: A partial or total gap of seasonal adjusted statistical indicators should be filled in by using, whenever possible, other official sources form national or international organisations. If the gap still persists, statistical indicators should be seasonally adjusted following as much as possible the general seasonal adjustment policy adopted at national level (if any) and agreed international standards and guidelines;

- Alternative B: The gap should be filled by seasonally adjusting statistical indicators by using default specification from international agreed methods such as the moving average based adjustment methods (U.S. Census Bureau) or the model based adjustment methods (Bank of Spain);

- Alternative C: Filing the gap by using not international agreed methods and software or not filing the gap at all.

22.3.4 Lack of history

Description

Statistical indicators, to be considered as candidates for the construction of cyclical composite indicators, should cover a sufficiently long time-period, lasting at least two or three cycles, meaning around 15 to 18 years at minimum. Unfortunately, official statistics are often subject to some recurrent shortening due to methodological changes, changes in standards, definitions, classifications etc. In some cases, statistical institutes provide adequately reconstructed consistent long time-series; even if they usually become available later with respect to the implemented changes. Additionally, some international institutions and central banks provide back re-calculated statistical indicators. Nevertheless, sometimes the compilers of composite indicators have to implement their own back-calculation either due to the unavailability of official long time-series or because available time series do not meet the main requirements for business cycle analysis (i.e. the past pattern should be kept as much as possible unchanged). In this specific case, very simple back-calculation methods can be appropriate. Such methods should be based either on chaining between old and new version of the same series over an overlapping period or on static regression models on the first log differences of both old and new versions of the series (assuming the existence of an overlapping period). Finally, some simple unobserved component back-calculation methods can be also applied.

Options

- Using as much as possible existing official back-recalculated series provided that they respect the principle of keeping the past unchanged;

- develop an own back-calculation strategy to be applied to statistical indicators which can be candidates in the construction of composite indicators, using a simple back-calculation methods and preserving as much as possible the past pattern of the indicators;

- Not considering the development of a back-calculation strategy even in absence of long official time-series.

Alternatives

- Alternative A: Develop a system of long time-series based on officially reconstructed ones (provided that they respect the principle of keeping the past pattern unchanged) and of own back-calculation based on simple statistical methods, based on the same principle;

- Alternative B: Only rely on an own back-calculation based on simple statistical methods respecting the principle of not changing the past pattern;

- Alternative C: Rely only on official back-calculation without considering their adequacy for business cycle analysis.

22.3.5 Variable selection

Description

Starting from the much larger set of macro-economic infra-annual statistical indicators, a variable selection exercise should identify the most suitable candidate series to be considered for the construction of a number of composite indicators. This first selection should be based on the ability of the series to accurately reproduce, anticipate or estimate in real-time the cyclical patterns observed in the economy, the turning points occurrence or the short-term patterns of preselected reference variables. This process should also allow identifying the leading/lagging characteristics of the indicators in the dataset. This exercise can be performed either by using parametric or non-parametric variable selection techniques, as well as graphical methods and be based on purely subjective judgment.

Options

- Parametric variable selection techniques;

- Non-parametric variable selection techniques;

- Graphical techniques;

- Preselection based on personal and subjective knowledge appreciation and skills about data;

- Combining several approaches;

Alternatives

- Alternative A: Combine parametric and non-parametric techniques with the support of graphical methods and of the experts' judgement to achieve the most efficient identification of the most suitable candidates to be included in the construction of cyclical composite indicators;

- Alternative B: Rely either on parametric or not -parametric variable selection techniques only;

- Alternative C: Rely purely only on graphical techniques or purely only on subjective judgement.

22.4 Composite indicators for estimating in real-time or forecasting cyclical movements

22.4.1 The reference variable

Description

The correct identification of the reference variable is crucial for this kind of composite cyclical indicators. In effect the reference variable defines the path and the movements which candidate variables are supposed to conform to, estimate in real-time or anticipate. Consequently, the variables which better reproduce the behaviour of the reference variable can enter as components series in a composite indicator of this type. Traditionally, the reference variable can be constituted either by a single statistical indicator, such as the GDP, or the industrial production index, or by a weighted combination of statistical indicators which is supposed to measure (according to The Conference Board approach and also the NBER original methodology) the unobserved economic conditions broadly defined many aspects of coincident economic activity.

The most representative reference variable for developed market economies is, as already mentioned, either the GDP or a latent variable such us the one defined by The Conference Board in its methodology. Also industrial production index can be still considered, even if its relevance is progressively declining as non-manufacturing or services sectors become more dominant in the composition of national economies. The main obstacle to the use of GDP is represented by its availability in almost all cases at quarterly frequency, often with publication delays, and its frequent and possibly large revisions.

By contrast, for emerging or developing economies it is possible that industrial production can still play a central role. For some developing economies or the underdeveloped ones, it is possible that reference variables should be more related to national peculiarities. Reference variables in this case could be related to energy production or price, mining activities, or even agricultural production. More complex is the identification of reference variables based on multiple indicators for such economies which require an in-depth investigation of specific features and characteristics of the national economy.

Options

- Using multi-indicators reference variables based on the characteristics and specificities of the national economy;

- Using a monthly GDP proxy, whenever available, or constructing it by means of distribution techniques;

- Using the industrial production index or a close proxy;

- Using specific national variables for developing and underdeveloped economies reflecting economic peculiarities and characteristics.

Alternatives

- Alternative A: For mature market economies, either the GDP (made available at monthly frequency) or a combination of indicators as in The Conference Board approach should be used as reference variables while the industrial production index should be used as an alternative or for emerging economies. For developing and underdeveloped economies, the most appropriate reference variable should be identified according to the characteristics of national economies, also taking into account its economic interpretability;

    – Alternative B: Use the industrial production index as the default reference series;

    – Alternative C: Do not specify any reference variable or use variables without a clear economic interpretation.

22.4.2 The identification of the reference cycle

Description

Originally, cyclical composite indicators have been developed to monitor the classical business cycle. Only later they have also been extended to analyze the growth cycle. Nowadays, composite indicators are also intended to monitor the acceleration cycle even if its high degree of volatility makes this task quite problematic and risky. Monitoring the classical business cycle or the growth cycle have both advantages and drawbacks: by concentrating on the former we can detect recessions but, during long expansionary phases this kind of indicator is not very informative about turning points; by contrast, by focusing on growth cycles we have more information about movements around the long-term trend even during expansionary phases, but we do not have any indication about risk or occurrence of recessions. Monitoring both cycles can provide a better picture of the cyclical evolution of the economy.

Options

- Construct composite indicators for the business cycle;

- Construct composite indicators for the growth cycle;

- Construct composite indicators for the acceleration cycle;

- Construct a couple of composite indicators for the classical business cycle and the growth cycle;

- Construct three composite indicators to monitor business, growth an acceleration cycle;

- Construct composite indicators to monitor other kinds of cycles such as the inventory cycle etc.

Alternatives

- Alternative A: Construct composite indicators to monitor both business and growth cycle with possible extensions to incorporate also the acceleration cycle;

- Alternative B: Construct a composite indicator for monitoring either the business or the growth cycle, being the choice based on the characteristics of past national cycles as well as on users' needs;

- Alternative C: Constructing composite indicators for monitoring either the acceleration cycle or other kind of cycles, such as short cycles (inventory cycles) or long ones (e.g. Kondratieff wave).

22.4.3 Detrending methods

Description

When constructing a composite indicator for monitoring the growth cycle, the reference variables as well as most of the component series must be subject to a filtering process to separate the trend from the cyclical component and estimate the deviation from the trend (i.e. the growth cycle). This additional filtering step can create further uncertainty, especially for the end of the sample and the current period. This is one of the reasons for which composite indicators focused on the growth cycle have been criticised in the past.

Several filtering techniques have been proposed since the development of the so-called phase average trend (PAT); These filtering methods are based on linear filtering techniques, on the decomposition of an ARIMA model, on state-space techniques or on multivariate extensions of the above mentioned methods. Some of the proposed methods, especially multivariate ones, imply some theoretical hypothesis which affect the shape of trend and cycle respectively, while others, especially univariate ones, are mostly data-driven. When constructing a composite indicator of the growth cycle, it is preferable to rely on simple, generally data-driven, methods and avoid constructing indicators influenced by one or another economic theory.

Options

- Using the HP filter;

- Using a finite-horizon approximation of the band pass filter;

- Using a two sided HP filter;

- Using a univariate or multivariate ARIMA decomposition such as the Beveridge and Nelson;

- Using a univariate or multivariate detrending method in the state space framework.

Alternatives

- Alternative A: Identify the best possible detrending solution by comparing within a simulation exercise (possibly in real-time) alternative univariate linear filtering techniques such as the single HP filter, the two sides HP filter and an approximation for a finite time-horizon of the band pass filter, using alterative specifications of the filters' parameters to better accommodate data characteristics, especially data volatility;

- Alternative B: Use default versions of either HP filter or the finite-horizon approximation of the band pass one;

- Alternative C: Use either non-standard linear filtering techniques or methods implying the acceptance of specific economic theories.

22.4.4 Estimation/aggregation methods

Description

The indicators proposed by The Conference Board (for the classical business cycle) and by the OECD (for the growth cycle) are widely disseminated and recognised. The two institutions already produce a large number of composite indicators for countries and regions and their methodology has been already applied by other national institutes in developing cyclical composite indicators. Both institutions use a relatively simple methodology to calculate the composite indicator as a weighted average of the component series where the weighting scheme based on some statistical criteria or on equal weights. Alternatively, the use of dynamic factor models has also gained a certain credibility and applications. Finally, other methods have also been proposed for constructing composite indicators to anticipate or estimate in real-time cyclical movements but they appear, at least for the moment, mostly academic exercises than solutions potentially applicable on a large scale. Furthermore, it has to be noted that using standard and simple techniques which are easily accessible by all countries can foster the cross-country comparability improving considerably the cyclical monitoring at global level.

Options

- Use of weighted averages of component series techniques such as those proposed by the Conference Board and the OECD;

- Use dynamic factor models based on either small or medium sized datasets;

- Use other linear or non-linear time series techniques;

- Use weighting schemes based on purely subjective appreciations.

Alternatives

- Alternative A: Identify the best performing method within a simulation exercise (possibly in real-time), comparing The Conference Board and the OECD approaches (for business and growth cycles respectively), customised to take into account data specificities of the country, and a few dynamic factor based composite indicators using a different size of the dataset;

- Alternative B: Using always one of The Conference Board or the OECD approaches;

- Alternative C: Using other linear or non-linear methods or purely subjective weighting schemes.

22.5 Composite indicators for detecting and forecasting turning points

22.5.1 The reference variable

Description

Turning point indicators aim essentially to estimate in real-time or to anticipate the occurrence of turning points without providing an estimation of the evolution of the cyclical components. For this reason, in this specific case, the identification of the reference variable is less relevant because we are interested only in its turning points and not in the shape of its cyclical components between turning points. Nevertheless, the reference variable should show a sequence of turning points which coincide with the turning points of the entire economy. Once again, the identification of the reference variable can vary quite a lot following the degree of development of various economies.

Options

- Using a multi-indicator reference variable;

- Using a monthly estimate of the GDP;

- Using a monthly industrial production index;

- Using country specific reference variable better describing the economy for developing and underdeveloped economies.

Alternatives

- Alternative A: For mature market economies, either GDP (estimated at a monthly frequency) or a combination of indicators such as the coincident indicators in The Conference Board methodology should be used as reference variables while the IPI should be used as an alternative or for emerging economies. For developing and underdeveloped economies, the most appropriate reference variable should be identified according to the characteristics of national economies, also taking into account its economic interpretability.

- Alternative B: Use the industrial production index as the default reference series;

- Alternative C: Do not specify any reference variable or use variables without a clear economic interpretation.

22.5.2 The identification of the reference cycle

Description

Monitoring the occurrence of turning points on multiple cycles is facilitated by the existence of unique sequences linking them. When looking at the business and growth cycles, there is the so called $ABCD$ sequence, where A and D are respectively the peak and trough of the growth cycle and C and D of the business cycle. Including also the acceleration cycle, the extended sequence takes the form $\alpha AB\beta CD$ where α and β are peak and trough of the acceleration cycle. These sequences offer an attractive way to monitor simultaneously the occurrence and sequence of turning points and consequently the various phases of the cycles.

Options

- Construct composite indicators for turning point detection of the business cycle;

- Construct turning points composite indicators for the growth cycle;

- Construct turning points composite indicators for the acceleration cycle;

- Construct turning points composite indicators for the business and growth cycle within the $ABCD$ framework

- Construct turning points composite indicators for the business, growth and acceleration cycles within the $\alpha AB\beta CD$ sequence;

- Construct turning points composite indicators using other definitions of cycles.

Alternatives

- Alternative A: Jointly monitoring the growth and the business cycle with the possibility also to include the acceleration cycle;

- Alternative B: Monitoring turning points of either the growth or the business cycle according to the characteristics of national cycles, taking into account also users' requests;

- Alternative C: Monitoring either the acceleration cycle or cycles based on other definitions.

22.5.3 Turning points historical dating

Description

The availability of historical dating (locating past turning points for a sufficiently long time-period) is of great importance when constructing turning points composite indicators and composite cyclical indicators. Historical dating will help in evaluating the performance of alternative models in identifying past turning points leading to the selection of the best performing model. Official turning point chronologies have been elaborated for a few countries and, whenever they exist, they should be used. In the absence of official chronologies, historical dating should be developed and regularly maintained. They should cover the cycles for which a turning point indicator is being constructed and they should be based, as much as possible, on the reference variables selected for constructing composite indicators.

Options

- Availability of official dating chronologies for one or more among business, growth and acceleration cycles;

- Constructing and maintaining historical turning point dating, ideally for business, growth and acceleration cycles or just for some of them, according to the turning points indicators to be developed;

- Irregularly maintained historical dating;

- Absence of any plan to construct historical dating.

Alternatives

- Alternative A: Develop and regularly maintain a system of historical dating covering the growth and the business cycle turning points and ,eventually, also the acceleration cycle turning points, using as much as possible all already available piece of information such as official chronologies;

- Alternative B: Develop an historical dating of turning points or using an existing one either for the business or for the growth cycle;

- Alternative C: Irregular maintenance of an implemented historical dating, absence of a historical dating.

22.5.4 The dating algorithms

Description

Dating rules have been implemented to allow a precise identification of past turning points. For this reason, they constitute the core of each historical chronology. Despite a certain proliferation of dating algorithms proposed in the literature, in applied business cycle analysis only two are widely used: the first was originated by Bry and Boschan (1971) and further extended by Harding and Pagan; the second originated by the work of Hamilton. The first dating rule is purely non-parametric, while the second is parametric, based on non-linear time-series modelling. The main advantage of the non-parametric dating rule based on the Bry and Boschan algorithm and its extension is the computational simplicity and the fact that it ensures that past turning points will remain unchanged even if new observations become available.

Options

- The Bry and Boschan dating algorithm with its extensions proposed by Harding and Pagan;

- The Hamilton dating rule;

- Other parametric and non-parametric dating rules.

Alternatives

- Alternative A: Using a non-parametric dating rule, such as the Bry and Boschan algorithm, with possible customisation to take into account peculiarities of each national economy;

- Alternative B: Using the parametric dating rule proposed by Hamilton adapting it to incorporate national peculiarities;

- Alternative C: Using other, less standard dating rules.

22.5.5 Methodology for the construction of turning points composite indicators

Description

Since turning points can be seen as discontinuities in a time-series and also considering that cyclical movements (especially of the classical business cycle) appear asymmetric, non-linear modelling techniques have been preferred for the construction of this kind of composite indicators. Markov switching models (MS) have been the most frequently used in this context, but also binary regressions (such as PROBIT and LOGIT) and other non-linear time-series modelling techniques, such as self-exciting threshold autoregressive models (SE-TAR) has received some attention. MS models can be used in a univariate setting, fitting a model to each component series of the composite indicator; in this case the final version of the composite indicator is obtained by averaging the signals returned by each univariate model by using a weighting scheme reflecting the reliability of each signal.

Alternatively, turning point composite indicators can also be developed in a multivariate way (MSVAR), fitting the model to all component series which eliminates the use of the weighting scheme described above. Furthermore, it has been proven that, within the ABCD framework, the use of a MSVAR model can allow to the simultaneous construction of a composite turning point indicator for the growth and the business cycle. For the SETAR models, the use of a multivariate version is more problematic due to computational features. Finally, binary regression models are used generally in the multivariate way. When constructing composite indicators using these techniques, it is important to pay particular attention to the correct model specification, especially in terms of number or regimes, presence/absence of heteroscedasticity, etc.

Options

- Using either univariate or multivariate MS models;

- Using SETAR models;

- Using binary regression models;

- Using some linear models.

Alternatives

- Alternative A: Identify the best MSVAR specification allowing for the joint construction of a pair of turning points composite indicators for business and growth cycle within a simulation exercise(possibly in real-time). Complement it with the identification of the best composite indicator for the acceleration cycle based on the fit of individual MS models to each components series;

- Alternative B: Find the best specification of composite indicators for the growth and business cycle independently computed either by using MS or binary regression models;

- Alternative C: Using other non-standard modelling techniques.

22.6 Growth composite indicators

22.6.1 The reference or target variable

Description

The aim of these composite indicators is to estimate in real-time or to forecast within a quite short horizon, the movement of some key macroeconomic indicators such as the GDP, inflation etc. In this sense, these composite indicators can be viewed as an alternative to traditional now/forecasting techniques to enhance the timeliness of key macroeconomic indicators. The reference variable, using the same terminology adopted in previous sections, is the one for which the present or the near future have to be estimated by the composite indicators.

For this reason, it could be more appropriate to call it a target variable instead. These composite indicators are consequently designed to replicate over a certain time-period and to anticipate the short-term movements of the target variables represented by its period on period or year on year growth rate (whenever the target series is not seasonally adjusted). In most cases, the target variables identified vary from the GDP, inflation, employment and, sometimes, industrial production. This is generally true for mature market economies and emerging economies while for developing and underdeveloped economies other variables may play a key role and then require the construction of composite indicators to anticipate their evolution.

Options

- Identify a small set of key macroeconomic indicators to be monitored for the present and the near future by means of growth composite indicators;

- estimating the present or future evolution only of a single variable, such as the GDP;

- Do not identify any variable to be either estimated in real-time or forecasted for the near future.

Alternatives

- Alternative A: Define a small set of macro-economic indicators to be estimated, by means of composite growth indicators, in real-time or forecasted over a short time-horizon to obtain a clear picture of the future economic situation;

- Alternative B: Using growth composite indicators to anticipate the movements of one or two indicators without pretending to obtain a picture of the economic situation in the near future;

- Alternative C: Do not construct any growth composite indicator.

22.6.2 The methodology

Description

In the construction of growth indicators, the methods more frequently used are based on dynamic factor models (based on either a small or a large dataset) or on regression models. Also VAR models have had some success. All these models have been extended in order to deal with the mixed frequency and the ragged edge structure of data, which is typical of this kind of composite indicators. Other, even more sophisticated methods have also been used, mostly in academic studies, but they do not have for now relevant applications in the regular construction and publication of growth composite indicators. Obviously, since the characteristics of each target variable are different from others, it is likely that the same model does not necessarily provide the best results for each target variable.

Options

- For each target variable, select the best performing model within a set of predefined models;

- Select a single model for all target variables within a set of predefined models;

- Using non-standard models.

Alternatives

- Alternative A: Starting from a set of models largely used in the construction of growth indicators, identify, for each target variable, within a simulation exercise (possibly conducted in real-time) the best performing model;

- Alternative B: Starting from a set of models largely used in the construction of growth composite indicators, identify which, within a simulation exercise (possibly conducted in real-time) which models provide the best results for all target variables;

- Alternative C: Using non-standard models for which there are not enough conclusive scientific or empirical evidence of their superior performance.

22.7 Revisions of cyclical composite indicators

22.7.1 Regular revisions

Description

All kinds of cyclical composite indicators are subject to regular revisions due to the revisions characterising most of their component series. Obviously, this kind of revision is not under the control of the compiler of cyclical composite indicators, since it is part of the so-called data production process. Another kind of revisions which can affect the behaviour of cyclical composite indicators is linked to the choices made concerning the re-estimation and re-specification of the models or of their weightings schemes. This kind of revision strongly depends on the decisions taken by compilers of cyclical composite indicators. They have to face the usual dilemma between the precision of their estimates, which suggests that composite indicators should be continuously re-specified and re-estimated to be as much as possible adherent to data; and the stability which suggests to keep the composite indicators as much as possible consistent over the time.

Options

- Re-specify and re-estimate the model or the weights each time;

- Re-specify the model each year and re-estimate parameters each time;

- Re-specify and re-estimate the model each year;

- Re-estimate and re-specify the model on an irregular basis;

- Never re-specify nor re-estimate the model.

Alternatives

– Alternative A: Review the model specification and re-estimate parameters each year unless the occurrence of unexpected events such as economic crises or structural changes suggest an intervention within the year;

– Alternative B: Review the model specification every year and re-estimate parameters each period;

– Alternative C: Never revise, revise too often or at an irregular basis.

22.7.2 Major revisions

Description

After a certain number of years, the chosen model specification or weights configuration for a given cyclical composite indicator can suffer of a certain degradation of their performance. Furthermore, statistical indicators can also be subject to big revisions due to the adoption of new definitions, methodologies and/or classifications. In such a context, an in-depth review of the cyclical composite indicators is necessary in order to avoid the proliferation of wrong or misleading signals. Usually, for main statistical indicators, a calendar for major revisions is made available by several countries. In this case, synchronising major revisions of cyclical composite indicators with those of main macroeconomic statistical indicators can minimise the impact of such large revisions.

Options

- revise cyclical composite indicators each time major revisions are scheduled for statistical indicators;
- revise cyclical composite indicators every 5 years;
- revise cyclical composite indicators every two to four years;
- never revise cyclical composite indicators;
- revise them on a purely irregular basis.

Alternatives

- Alternative A: Perform a comprehensive revision of model specification, weights, definitions, and data selection process of cyclical composite indicators whenever major revisions of statistical indicators take place and every five years. Whenever possible, major revisions in underlying components and methodologies should be grouped together, incorporated, and announced at the same time to minimize disruptions;
- Alternative B: Conduct an in-depth revision of cyclical composite indicators every five years;
- Alternative C: Do not conduct any in-depth revision of cyclical composite indicators or do it irregularly.

Handbook on Cyclical Composite Indicators

FOR BUSINESS CYCLE ANALYSIS

The 2007-2009 global financial and economic crises revealed severe weaknesses in the system of macroeconomic infra-annual statistics which prevent a prompt detection of the crises especially when in the making.

The purpose of this handbook is to provide statistical and econometric guidance on harmonized principles for the compilation and monitoring of cyclical composite indicators. At the same time, the handbook outlines standards for the compilation and presentation of such indicators.

For more information
http://ec.europa.eu/eurostat/

Publications Office