Handbook on Rapid Estimates

2017 edition

United Nations

MANUALS AND GUIDELINES

eurostat

Handbook on Rapid Estimates

2017 edition

Manuscript completed in July 2017

Luxembourg: Publications Office of the European Union, 2017

PRINT: ISBN 978-92-79-68896-6 ISSN 2363-197X doi: 10.2785/544610 Cat. No: KS-GQ-17-008-EN-C
PDF: ISBN 978-92-79-66897-3 ISSN 2315-0815 doi: 10.2785/4887400 Cat. No: KS-GQ-17-008-EN-N

Foreword

The 2007-2009 global financial and economic crisis led to a reflection on the need for an international agreed system of high frequency macroeconomic statistics and indicators. At the time it was recognised that high frequency statistics and indicators based on an international agreed methodology could facilitate the early detection of changes in macroeconomic conditions. Providing relevant accurate reliable and timely data, e.g. based on so-called rapid estimates, is therefore of the essence to facilitate the monitoring and assessment of policies.

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

During the discussions at the seminars, it emerged that—on the one hand—there were large differences across countries in the timeliness of key macroeconomic indicators. On the other hand, a consensus emerged that guidance for the compilation of rapid estimates should be prepared based on best international practices. In addition, it was deemed necessary to clarify the terminology associated with rapid estimates. In order to establish a common understanding of rapid estimates, Eurostat took the lead in drafting this handbook and preparing of a glossary of terms for rapid estimates to clarify the different typologies of rapid estimates, their purposes and characteristics.

The handbook presented herewith outlines practical and suitable compilation methods for rapid estimates. It draws on a wide range of experience and expertise and benefits from recent theoretical and practical developments in the area. The handbook is intended to assist those producing rapid estimates e.g. in the area of key short term macroeconomic indicators. It is also intended to assist countries that plan to set up a more comprehensive system of rapid estimates by providing both methodological foundations for their compilation and by giving practical guidance on individual steps and elements of the underlying compilation process.

This handbook should be considered as both a reference tool, stating the state of the art in the area of rapid estimates and a guide towards the implementation of rapid estimate systems in organisations. It has been designed to meet the requirements of a wide audience, both technical and non-technical, be it academics, research bodies, private institutions or Government entities. It is therefore an invaluable tool and highly recommended for anyone wanted to develop a better understanding of rapid estimates.

The Handbook benefited from comments from experts who participated in the Expert Review, namely (in alphabetical order of countries followed by international organizations): Zsuzsanna Szõkéné Boros (Hungarian Central Statistical Office, Hungary), Luciana Crosilla and Solange Leproux (Italian National Statistical Institute - ISTAT, Italy), Luca Farnia (Fondazione Eni Enrico Mattei; Centro Euro-Mediterraneo per i Cambiamenti Climatici, Italy), Marco Gallegati (Università Politecnica delle Marche, Italy), Filippo Moauro (Italian National Statistical Institute - ISTAT, Italy), Ieva Ainare (Central Statistical Bureau of Latvia, Latvia), Ma. De Lourdes Mosqueda González (Instituto Nacional de Estadística y Geografía - INEGI, México), Klaus Abberger (KOF Swiss Economic Institute at the ETH Zürich, Switzerland), Jennifer L. Castle, David F. Hendry and Oleg I. Kitov (Oxford University, United Kingdom), Pawel M. Stano (European Commission Joint Research Centre, JRC).

The Handbook also benefitted from comments and suggestions from national statistical offices (NSI), central banks, regional commissions and international organizations during the global consultation that took place during the period in December 2016 to February 2017.

The publishing institutions would like to express their thanks and gratitude to the editor of this handbook, Gian Luigi Mazzi (formerly: Eurostat), and his co-editor, Dominique Ladiray (French National Institute of Statistics and Economic Studies, INSEE, France), for the time and effort they dedicated to bringing this handbook to fruition. Also thanks to all the experts who have contributed to the drafting of the various chapters in this handbook. Thanks also the support and comments from Mr. Ivo Havinga, Mr. Herman Smith and Ms. Ilaria Di Matteo (UNSD) who facilitated the wider consultation of the various drafts of the document and for their invaluable inputs.

A special thanks goes to Dan A. Rieser, Eurostat, for his role as technical editor and the supervision of the entire publication process.

Editors and Authors

Editors

Editor:	Gian Luigi Mazzi (formerly: Eurostat)
Co-Editor:	Dominique Ladiray (INSEE)
Technical Editor:	Dan A. Rieser (Eurostat)

Contributing Authors (in alphabetical order)

Fabio Bacchini
Istat
bacchini@istat.it

Karim Barhoumi
IMF
kbarhoumi@imf.org

Dario Buono
Eurostat
dario.buono@ec.europa.eu

Jennifer Castle
Magdalen College Oxford
jennifer.castle@magd.ox.ac.uk

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

Olivier Darné
Université de Nantes
olivier.darne@univ-nantes.fr

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

Claudia Foroni
European Central Bank
foroni.claudia@gmail.com

Riccardo Gatto
Eurostat
riccardo.gatto@ec.europa.eu

David Hendry
Nuffield College Oxford
david.hendry@economics.ox.ac.uk

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

Oleg Kitov
St Hilda's College Oxford
oleg.kitov@economics.ox.ac.uk

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

Silvia Lui
University of Groningen
s.w.s.lui@rug.nl

Editors and Authors

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

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

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

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

Katerina Petrova
University of St. Andrews
Katerina.Petrova@st-andrews.ac.uk

Tommaso Proietti
University of Rome Tor Vergata
tommaso.proietti@uniroma2.it

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

Table of contents

1 Introduction: Objectives, Definitions, Costs and Benefits of Rapid Estimates

Authors

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

1.1 Introduction

The 2007-2009 global financial and economic crisis, the so-called Great Recession, revealed some weaknesses of the system of macroeconomic infra-annual statistics which prevent a prompt detection of the crisis signals. Furthermore, they could not help policy makers to quickly react in order to mitigate most of the crisis effects. Obviously, the situation of infra-annual statistics was quite different across countries but some weaknesses were, in a more or less pronounced way, common to all of them. As stated in European Union and United Nations (2017), the main observed weaknesses affecting the system of infra-annual macroeconomic statistics can be synthesized as follows:

- the lack of a homogeneous set of statistical indicators able to provide a reliable overall picture of the macroeconomic situation and allowing cross-country and region comparisons;
- the lack of timeliness observed for most of key macroeconomic indicators and in particular for the quarterly national accounts, especially GDP;
- the difficulties encountered by users and analysts in extracting proper cyclical signals directly from official statistics; and
- the lack of new cyclical indicators (maybe composite ones) providing a clear picture of cyclical movements and delivering reliable and timely messages on turning points occurrence.

Starting from limits of official statistics observed during the great recession, the United Nations Statistical Division (UNSD), the statistical office of the European Commission (Eurostat) and the Center 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 organization 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, academics and users took part to these seminars, 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 discussion so that the main weaknesses were identified.

Based on the outcome of these international seminars, the United Nations Statistics Division and Eurostat, in close collaboration with Statistics Canada, Statistics Netherlands and the 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 four thematic areas, namely:

- **Rapid estimates**: The programme of work on rapid estimates included the preparation of a glossary of rapid estimates in order to achieve a consensus around a common terminology associated to various kinds of estimates produced by statistical agencies or other private or public bodies; and the preparation of a handbook on rapid estimates providing a methodological overview of statistical and econometric techniques useful to foster timeliness of macroeconomic statistics as well as some guidance on the construction of rapid estimates;
- **Business cycle composite indicators**: The work programme of work on business cycle composite indicators included the preparation of a handbook on cyclical composite indicators (see European Union and United Nations (2017)) presenting the various kinds of composite indicators and the associated statistical and econometric compilation techniques as well as some guidance for their construction;
- **Economic tendency surveys**: The programme of work included the preparation of a handbook on economic tendency surveys (see United Nations (2015)) to provide internationally harmonized principles to be used in the conduct of tendency surveys built on the existing work undertaken by countries

and international and regional organizations, most notably the work carried out as part of the Joint Harmonized European Union Programme of Business and Consumer Surveys and the related guidelines issued by the European Commission Directorate General for Economic and Financial Affairs and the work of OECD

- **Data template for short term indicators and analytical indicators**: The programme of work included the development of an internationally accepted list of indicators for short-term economic statistics suitable for macroeconomic surveillance, early warning of economic and financial vulnerabilities and detection of turning points in business cycles, and the preparation of a handbook (see United Nations (2016)) providing methodological reference and practical guidance for those indicators together with their metadata.

The overall objective of those initiatives was to enhance the set of available tools for effective monitoring that could eventually be implemented as a real-time economic and cyclical early warning system. The idea of an effective monitoring of the short term macroeconomic situation was presented in Mazzi et al. (2013) and an attempt to design a real-time monitoring system was proposed by Mazzi et al. (2016).

The present handbook, jointly prepared by Eurostat and the United Nations Statistical Division, aims to provide an answer to the first two corrective measures listed above, namely the preparation of an handbook on rapid estimates and of a glossary for rapid estimates. It is worth noticing that this handbook does not aim to influence political decisions of statistical institutions about publishing or not rapid estimates; instead, it intends to clarify typologies of rapid estimates, to conduct a wide review or methodologies for the enhancement of timeliness and to provide compilation guidelines.

The structure of this chapter is as follow:

- Section 1.2 briefly introduces some aspects related to the improvement of timeliness and the use of rapid estimates;
- Section 1.3 introduces the aim, scope and limitations of this handbook;
- Section 1.4 provides a short overview of the structure of the handbook and the main content of each part while
- Section 1.5 provides some general conclusions.

1.2 Timeliness improvement and rapid estimates

Statistical agencies, central banks and other institutions active in the statistical field, are committed to produce and publish high quality statistical information. Timeliness is one dimension of data quality and producers of statistics must take care of it. Unfortunately timeliness is often in competition with other dimensions of data quality such as accuracy and reliability . This imply that, not necessary, all dimensions of data quality can be maximized simultaneously. Consequently priorities among data quality dimensions have to be identified. For quite a long time, especially in Europe, precision and reliability of statistics have received higher priority with respect to timeliness (see chapter 4). The situation in Europe started to change in the second half of the eighties and more significantly in the nineties mainly due to the process of economic and monetary unification. However if looking at the production of key macroeconomic statistics, it appears that the attention paid to timeliness is still very different across countries and regions. The reason for this persisting gap can certainly be attributed to the differences in the development of statistical systems among countries and regions.

Various strategies can be used when aiming to improve timeliness, some of them fully in line with the traditional ways of producing official statistics, others trying to combine traditional methods with more advanced statistical and econometric ones. A short discussion of various strategies for timeliness enhancement is presented in chapter 2. When considering the possibility of combining traditional statistical production with more advanced statistical and econometric methods, the role of the latest can vary quite a lot across different kind of

rapid estimates (see section 2.B). The attitude of statistical agencies regarding the incorporation of advanced statistical and econometric techniques into the official statistics production process can differ quite a lot.

Some statistical agencies tend to be more conservative, willing to incorporate only a minimum contribution of non standard statistical and econometric techniques. They in general prefer to invest in speeding up the statistical production process and in its modernization. Others countries may be more open to invest in combining traditional methods and advanced ones. Obviously both positions have their advantages and drawbacks. We can say that, in general, speeding up the statistical production as an instrument to improve the timeliness has the advantage that new figures are fully consistent with the past ones. On the other hand, since the speeding up of the statistical production process or the modernization of the statistical system are processes scheduled over a medium- and long-term horizon, if there is a need for improving timeliness in the short term, then the use of advanced statistical and econometric techniques can provide a transitory solution waiting for the modernization process to take place.

Being realistic, it also appears that the modernization of the statistical system and the speeding up of the statistical production process can lead to limited gains in timeliness and will not be sufficient to estimate in real-time or almost in real-time the trend of some key macroeconomic indicators. To achieve such an objective, a change of paradigm might also be necessary. One idea could be to release the link between the reporting period and the reference period. For example, if during the third week of month t, you ask businesses to report on their production of the 2 first weeks of the same month t, this will automatically produce gain of 14 days in terms of timeliness which, even by modernizing the production system without using any statistical or econometric techniques, can lead to obtain almost real-time estimates for the key macroeconomic variables.

These considerations are based on the actual state of art concerning data acquisition and data processing. But the data revolution and the rapid technological and IT advances can change completely this scenario. Unfortunately, despite the growing interest in the use of “big data”, we are not yet in the position of formulating any realistic hypothesis concerning future scenarios in this specific context. Chapter 17 gives more details on the usefulness of big data especially for rapid estimates.

1.3 Aim, scope and limitations of the handbook

This handbook aims to provide an overview of various types of rapid estimates together with a clear set of definitions proposed to avoid any kind of confusion in terminology. It describes in details the most commonly used statistical and econometric techniques for constructing various kind of rapid estimates. 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 formalizations 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 presentation of statistical and econometric tools useful for the construction of rapid estimates so that other kinds of estimation techniques, such as these for missing data, are not discussed here. The use of composite indicators to estimate in real-time or anticipate the trend of some key macroeconomic variables is also not presented here. It is, by contrast, mainly presented in the Handbook of Cyclical Composite Indicators (European Union and United Nations (2017)) which dedicates to them a complete part.[1] We suggest readers interested in this topic should refer to the aforementioned chapters of the Handbook on Cyclical Composite Indicators (European Union and United Nations (2017)) mentioned in the footnote.

The handbook also presents selected readings about well assessed and consolidated methodologies, so that most recent developments, still in an experimental phase or restricted to the academic world are not

[1] In this part, various kind of composite indicators to enhancing timeliness are presented by Mazzi and Cales (chapter 15), Camba-Mendez et al. (Chapter 17), Luciani (chapter 16) and Charpin (chapter 18) while in Mazzi and Ruggeri-Cannata (chapter 21) an empirical comparison of some coincident indicators to anticipate GDP movements is presented.

considered. In addition, structural models which imply the adoption of one or another economic theory are also not considered. The focus in the handbook is mainly on data driven methods or on methods based on statistical evidence. Nevertheless, producers of rapid estimates should pay special attention on the fact that their figures should be plausible both from the statistical point of view and from the economic theory point of view, avoiding the use of any implausible relationship. The target readers of the handbook are mainly those compiling rapid estimates within statistical offices and central banks. Also researchers, students and academics who can benefit from the contents of this handbook.

1.4 Structure of the handbook

Part I "**General Aspects**" presents some general aspects of rapid estimates. In particular, Chapter 2 discusses alternative ways to enhance timeliness and different kind of rapid estimates. In the annex, the glossary of rapid estimates as developed by Eurostat (see Barcellan et al. (2009)) is presented. Chapter 3 provides a very detailed formalization of the nowcasting and forecasting concepts and methods [2]. Chapter 4 presents some considerations on the trade-off between timeliness and accuracy.

Part II "**Statistical and Econometric Techniques for Rapid Estimates**" describes basic modeling aspects. Chapter 5 provides a wide overview of modeling techniques for rapid estimates starting from the univariate ones until the most advanced multivariate ones. Chapters 6 and 7 mainly deal with variable and model selection techniques.

Part III "**Rapid Estimates based on Data Available at Different Frequencies**" is devoted to methods able to deal with data available at different time frequencies. Chapter 8 presents temporal disaggregation techniques in a new and very attractive way while chapter 9 provides an overview of mixed frequency modeling techniques.

Part IV "**Rapid Estimates based on Factor Models**" describe some aspects related to the use of factor models for rapid estimates. Chapter 10 is devoted to an in-depth analysis of factor models especially dynamic ones. Since this chapter is a reprinting of a paper already published, it covers a wider range of applications than just rapid estimates. Chapter 11 compares the performances of small, medium and large size dynamic factor models in the nowcasting exercise.

Part V "**Rapid Estimates based on Combining Forecasting Techniques**" deals with the use of combining forecasting techniques in constructing rapid estimates. Chapters 12, 13 and 14 analyze in details both from the theoretical and the empirical points of view the usefulness of combining forecasting techniques in the specific context or building up rapid estimates.

Part VI "**Aggregate versus Disaggregate Approaches to Constructing Rapid Estimates**" deals with the issue of using aggregated or disaggregated data in the construction of rapid estimates. Chapter 15 shows the relative performance of rapid estimates built up by using geographically aggregated information only; indirect one and both kind of information in a mixed model.

Part VII "**Big Data and Rapid Estimates**" introduces the role that big data could play in constructing rapid estimates. Chapter 16 describes some challenges related to the use of big data in rapid estimates while chapter 17 contains some guidance for the use of big data in rapid estimates models.

Part VIII "**Constructing and Evaluating Rapid Estimates**" analyzes the ways in which rapid estimates should be evaluated in chapter 18 and presents guidelines on the construction of rapid estimates in chapter 19.

[2] This chapter for its complexity appears to be more oriented to experienced users.

1.5 Conclusions

This introductory chapter presents the historical background which led to the decision of preparing this handbook and the approach followed in its preparation. There is no doubt that rapid estimates can constitute a very useful component of the information provided by traditional sources such as official statistics. Rapid estimates can contribute to improving timeliness of macroeconomic variables but also to provide anticipation of short-term movements for the current period or for the near future of key macroeconomic indicators.

In this respect, rapid estimates are an essential component of a real-time economic monitoring. Such a system 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 to prevent all new crises, more timely information on key macroeconomic variables can 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 cost.

In a globalized world, the availability of comparable rapid estimates across countries and regions becomes a further element in strengthening the implementation of quick and effective counter-cyclical measures, coordinated across countries.

Bibliography

Barcellan R., Mazzi G.L., Montana G. and Ruggeri Cannata R. (2009). Glossary of statistical terms associated to the estimation of key economic indicators, International Seminar on Early Warning and Business Cycle Indicators, Scheveningen, The Netherlands. December 2009.

European Union and United Nations (2017), Handbook on Cyclical Composite Indicators, edited by G. L. Mazzi, co-edited by A. Ozyildirim, European Union and the United Nations, Luxembourg, 2017.

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., 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.

United Nations (2015), Handbook on Economic Tendency Surveys. Statistical papers, Series M No. 96. https://unstats.un.org/unsd/nationalaccount/docs/ETS_Handbook_Final.pdf.

United Nations (2016), Handbook on Data Template and Metadata for Short-Term Statistics. Studies in methods Series F, No. 117. https://unstats.un.org/unsd/nationalaccount/docs/Data_Template.pdf.

General Aspects

2 Rapid Estimates: Different Products for Different Purposes

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

2.1 Introduction

Official statistics are used as a daily reference and source of information to interpret how economic and social phenomena are evolving. They also constitute the basis in the decision and policy-making process. Users, especially policy and decision makers as well as business cycle analysts, require more and more timely statistical information in order to increase their ability to monitor the economic situation and to design effective economic and monetary policies. As a matter of fact, increasing timeliness of regular estimates is not easily achievable in the short-term. For this reason several statistical offices and central banks as well as other statistical institutions and research institutes have considered new ways of producing more timely estimates of key macroeconomic indicators. In this effort, they have investigated the possibility of integrating and complementing the traditional production process by advanced statistical and econometric techniques. The consequence of this activity by statistical bodies has been the more timely release of macroeconomic indicators based on a variety of new products which are commonly grouped under the umbrella of rapid estimates. Such new products differ, among others, by the intensity degree of the utilisation of statistical and econometric techniques.

Unfortunately, a certain lack of harmonisation in labelling all these new statistical products has been observed. In the international discussion fora, starting with the series of seminars jointly organised by UNSD and Eurostat as a follow up of the global financial and economic crises, the need of defining a common terminology for labelling the various types of rapid estimates has been clearly raised and put as a priority for the working activity of international organisations.

The need of such terminological harmonisation has been mentioned in Barcellan et al. (2009). In that paper the authors launched the initiative of constructing a glossary on rapid estimates which has been then compiled by Eurostat (see section 2.B). The purpose of this chapter is to provide a less formal description of some of the most commonly used kind of rapid estimates by synthetically sketching their main features. We aim then to provide a rationale to the disambiguation of a variety of rapid estimates, based on the identification of key elements which should be taken into account when characterising them. We leave to the glossary the role of providing a more comprehensive, formal and technical definition of the various products labelled under rapid estimates.

Following this introduction, section 2.2, will start discussing the relevance of a common terminology while section 2.3 will identify some key elements which have to be considered when characterising rapid estimates. Section 2.4 will shortly present a first definition of some types of rapid estimates based on the elements presented in section 2.3, section 2.5 introduces the possibility of developing higher frequency indicators as an indirect way to improve timeliness; section 2.6 analyses the various types of rapid estimates in terms of the variety of needs stemming from the users community while section 2.7 concludes. A list of the most recurrent terms related to the compilation of rapid estimates is annexed to this chapter for readers' convenience (2.A).

2.2 Relevance of a common terminology

Several statistical glossaries are available in the literature or on-line, such as the OECD Glossary of statistical terms and others available on university websites. Nevertheless, terms such as "nowcasting" or "flash estimate" are rarely included and there is a lack of consensus on the characteristics and boundaries of the statistical concepts covered. A qualification of statistical terms associated to rapid estimates of key economic indicators would undoubtedly foster a common understanding facilitating the discussion and exchange of opinions among official statisticians and users.

Although this could seem a minor issue with respect to more technical aspects involved in the construction of rapid estimates, it has to be stressed that various types of rapid estimates can be characterised by different degrees of reliability associated to their timeliness, to the completeness of the information sets used, etc. The lack of consensus in labelling them can affect the transparency and the perception of rapid estimates and create confusion among users. Furthermore, this can have negative effects on the comparability among

countries and statistical areas. The misinterpretation of rapid estimates released by statistical offices can also affect the effectiveness of decision and policy making processes and, as a side effect, hamper the credibility of official statistics.

The common terminology should be simple, precise, operational and unambiguous. The related glossary is intended as a simple list of definitions that clearly identify the typology of rapid estimates within a well-defined conceptual framework. The proposed terminology and the related glossary should be seen as evolving over the time in order to incorporate new developments in the family of rapid estimates. Finally, even if the common terminology is stemming from the area of official statistic and it is targeted to its main users, the consensus of the scientific and academic world has also to be actively pursued. In this way, it could become a reference for all actors working with official statistics. This first proposal for a statistical terminology could be progressively extended by an additional information layer dealing with a more in-depth analysis of methodological aspects, enriching the underlying conceptual framework.

2.3 Characterising rapid estimates

Rapid estimates refer to a timely numerical evaluation (i.e. an estimate) of an economic variable of interest, using the real-time data flow, in such a way that an early picture of the present, the recent past or the near future can be formed. Rapid estimates therefore include different types of estimates such as nowcasting, forecasting, flash estimates, leading and coincident indicators etc. These estimates however are not uniquely and clearly defined. Below is a description of the elements that are necessary to clearly and uniquely characterize different types of rapid estimates.

The starting point for the reflection on a common terminology for rapid estimates has been the discussion on statistical terms originated in the International Seminar on Timeliness, Methodology and Comparability of Rapid Estimates of Economic Trends (Ottawa 2009)[1], where no unanimous consensus was found on what terms like statistic, nowcasting, forecasting, flash estimate, leading and coincident indicator stood for (see Barcellan et al. (2009)).

The target of the common terminology is to identify and qualify the estimation (= assessment) of a measure of a given short-term economic indicator for a given reference period in a specific point in time.

In doing so, the following characterising elements should be taken into account:

- the relative position in **time** of the estimation with respect to the reference period;
- the **information set** used for the estimation;
- the statistical **method** applied for the estimation;
- the adherence of the estimation process to the regular **production process**;
- the **target variable** to be assessed.

The interrelations of these dimensions result in a matrix whose cells will correspond to the statistical terms and concepts covered by the key terms of the glossary. Whilst, in principle, the same importance could be attached to all these elements, in practice, in defining the different kind of estimates of economic indicators, following Barcellan et al. (2009), priority is given to specific elements.

In the proposed approach, time plays a prominent role, followed by the information set and then by the method and the adherence to the production process (at the same level); finally, the target variable follows.

a. The time dimension: statistics vs. nowcasts vs. forecasts

Let's focus on the estimation of a measure of a given short-term economic indicator y for a given reference period t in a specific point in time r: $y_{t,r}$ Then, with respect to time:

[1] Report on the seminar in Ottawa: http://unstats.un.org/unsd/nationalaccount/workshops/2009/ottawa/AC188-5.PDF

- if $t < r$ i.e. we are attempting to estimate a past measure of y, then we refer to the estimation as a statistic;
- if $t > r$ i.e. we are attempting to estimate a future measure of y, then we refer to the estimation as a forecast;
- if $t \approx r$ i.e. we are attempting to estimate a contemporaneous measure of y (just before or right after the end of the reference period t), then we refer to the estimation as a nowcast.

However, the estimation (statistic, nowcast or forecast) need to be further qualified by the other elements because this first characterisation is not enough to uniquely identify estimates. We will then analyse the definition of flash estimate focusing on the concepts needed to uniquely identify different estimates. Let's assume the following general definition of flash estimate.

Example 1.a: Definition of flash estimate

Flash estimate: ***An early estimate produced and published for an economic variable of interest over the most recent reference period.*** The flash estimate is normally calculated on the basis of an incomplete information set, however produced using almost the same statistical or econometric models as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.a:

A nowcast is a contemporaneous estimate of y just before or right after the end of the reference period t.

b. Information set

A key element to qualify an estimate is the information set used to produce it. The information set corresponds to the input to the estimation process and broadly speaking, in this field of investigation (economic indicators), it corresponds to:

- qualitative (soft) information;
- contemporaneous quantitative (hard) information;
- past quantitative (hard) information.

The qualification of the information used is clearly related to the other elements. It might depend on, for example, when the estimation is run, the composition of the information set and its time coverage. Forecasts, nowcasts and statistics are all based on the combination, to a greater or lesser extent, of the above mentioned input. The nuances among the different estimates result, therefore, from the different mix of information in a specific point in time.

Example 1.b: Definition of flash estimate

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. ***The flash estimate is normally calculated on the basis of an incomplete information set***, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models ***should incorporate hard data as much as possible***.

Example 2.b:

(i) A nowcast is a contemporaneous estimate of y just before or right after the end of the reference period t.

(ii) A nowcast is usually based on an information set composed by a mix of qualitative (soft), contemporaneous and past quantitative (hard) information.

Again the first two elements alone are not enough to uniquely qualify a nowcast. A coincident indicator would also satisfy these qualifications.

c. Statistical methods

Different statistical/econometric methods may be applied to produce an estimate (statistics, nowcasts or forecasts). Several families of methods are available for the estimation of key economic indicators. The choice between them depends on the combination of the other qualifying elements and on the available techniques and models. The time element and the available information play a prominent role in the choice. The methods range from those more accepted and spread in the production of official statistics (traditional grossing-up, sampling, combination of existing information within the business architecture of statistical authorities) to more statistically and/or econometric oriented techniques (such as extrapolation, structural models, factor models, etc.).

Statistical and econometric methods, combined with the other qualifying elements, determine also the statistical models applied to derive the estimation of the target variable. It is worth to note that, whilst considerations associated to economic modelling should be avoided in setting up a statistical model and the related statistical method(s), in practice, implicit economic assumptions underline the compilation of official estimations of key economic indicators (e.g., the concept of GDP and its decomposition in components is directly derived from the neo-Keynesian theory).

Example 1.c: Definition of flash estimate

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of an incomplete information set, however produced using the same ***statistical or econometric model*** as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Example 2.c:

(i) A nowcast is a contemporaneous estimate of y just before or right after the end of the reference period t.

(ii) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

(iii) A nowcast relies on specific dedicated statistical methods that optimise the use of the available information.

We are still unable to uniquely identify a nowcast with the elements specified so far, as a coincident indicator would also be covered by these qualifications.

d. Adherence to the regular production process

In the field of official statistics, particular relevance assumes the regular production process that reflects the business architecture in place in National Statistical Institutes or other statistical authorities. The regular production process corresponds to the "normal" way of producing official statistics on the basis of the available information. In particular, in the case of key economic indicators, it embeds the way of deriving macro-economic aggregates, usually at national level, from the basic information composed by administrative sources, results from surveys, etc., as well as the accounting constraints (if any) generated by the framework of the compilation (e.g. the National Accounts framework for GDP).

The regular production process covers also additional characteristics such as revision and release policy that help to further qualify the estimation. Therefore, terms like advanced, preliminary, first, second, regular, final estimate are derived from these additional characteristics.

Example 1.d: Definition of flash estimate

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of an incomplete information set, ***however produced using the same statistical or econometric model as for regular estimates***. Flash estimate models should incorporate hard data as much as possible.

Example 2.d:

(i) A nowcast is a contemporaneous estimate of y just before or right after the end of the reference period t.

(ii) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

(iii) A nowcast relies on specific dedicated statistical methods that optimise the use of the available information.

(iv) The information set and the methods used in producing a nowcast may differ from the methods applied in the regular production process of the target economic indicator.

e. Target variable: indicators vs. proxies

Sometimes the direct estimation of the latest value of the target variable is not feasible or it is technically too complex. Therefore, in order to give an appreciation of the recent or future evolution of the target variable over time, the object of the estimation is not the variable itself but a proxy that well approximates the characteristics and the behaviour in time of the target variable. In this context the relevant aspects to look at are the properties of the proxies, for example their leading, lagging or coincident evolution with respect to the target variable. This will permit us to characterise lagging, coincident or leading indicators.

Example 1.e: Definition of flash estimate

Flash estimate: An early estimate produced and published for ***an economic variable of interest*** over the most recent reference period. The flash estimate is normally calculated on the basis of an incomplete information set, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

We are now able to uniquely identify a nowcast thanks to the qualifications introduced in i-v.

Example 2.e:

(i) A nowcast is a contemporaneous estimate of y just before or right after the end of the reference period t.

(ii) A nowcast is usually based on an information set composed by a mix of qualitative, contemporaneous and past quantitative information.

(iii) A nowcast relies on specific dedicated statistical methods that optimise the use of the available information.

(iv) The information set and the methods used in producing a nowcast may differ from the methods applied in the regular production process of the target economic indicator.

(v) A coincident indicator is a contemporaneous estimate (nowcast) of a proxy of y.

2.4 Further considerations on rapid estimate types

Based on the elements described in section 2.3, we can now try to qualify some of the most commonly used rapid estimates by providing some more details about their characteristics. Such characteristics will constitute an important element when developing formal definitions in annex 2.B.

2.4.1 Extrapolation

Strictly speaking in mathematical terms, extrapolation is the process of constructing new data points outside a discrete set of known data points. In other words, extrapolation techniques allow the computation of latest and

future values of a time series starting from the already existing values. Extrapolation techniques are purely univariate, so that new values are computed on the basis of the past history. Extrapolation techniques can be either parametric or non-parametric, deterministic or stochastic, linear or non-linear. Those techniques vary from the simple curve extrapolation (linear, polynomial and logistic, etc.) to the smoothing ones such as exponential smoothing, moving average, linear time series models (ARIMA, unobserved components models), and to the non-linear time series models (ARCH, GARCH, TAR, STAR, SETAR, etc.).

Extrapolation techniques can be applied either directly on the reference variable or indirectly on its components. In the second case, new values of the reference variable will be derived by aggregation of the values computed for the components, taking into account the breakdown structure of the target variable and the related aggregation constraints. Extrapolation techniques can be used before, during and after the reference period and the time horizon can cover 1, 2 or even more periods ahead taking into account that the quality of results will progressively worsen with the length of the horizon. The implicit assumption when using extrapolation is that the recent past, the present and the future evolve with the same features of the observed past. As a consequence, it can work very well in periods characterized by regular evolution but it will be, by definition, unable to detect turning points, regimes changes or structural breaks.

Despite the limits described above, extrapolation techniques can be very useful when there is a need to compute recent or future values but no additional information is available on the reference variable. Extrapolation techniques can also be included in more complex rapid estimation processes, such as flash estimates, for example to compute values for geographical or sectoral components of the target variable (possibly of limited weight in the total) for which there is no additional source of information available.

Extrapolation techniques are widely accepted and implemented because they give quite robust results in the short term, especially one step ahead; moreover, it is not easy to construct estimators beating univariate extrapolation models (see Chapter 5) and they are then commonly used as benchmarking for more sophisticated models.

2.4.2 Nowcasting

In general terms nowcasting is the real-time or quasi real-time estimation of the evolution of a given variable. More specifically, nowcasting allows estimating the latest or the current movements of a reference variable. Nowcasting can take place, right before or after the end of the reference period. For example, if the last available information of the reference variable refers to the period $t-1$, its nowcasting for the period t should take place a few days before the end of the period t or few days after the beginning of the period $t+1$. Nowcasting is usually based on an incomplete information set with respect to the one used for the normal compilation process. The information set can be incomplete either from a geographical or a sectoral point of view, as well as in temporal terms: e.g. for a quarterly reference variable, only two months out of three of the related variables included in the information set are available. Concerning the type of the indicators included in the information set they can be:

- partial information on the reference variable (in geographical, sectoral or temporal way);
- qualitative data of the same sector and nature of the reference variable;
- other quantitative variables, statistically and logically related to the reference variable.

The information set should not contain variables which implicitly indicate the acceptance of a specific economic theory.

Nowcasting techniques can cover a large variety of multivariate methods such us: regression methods, multivariate time series methods (e.g. VAR), static and dynamic factors models, principal components methods, etc. In order to increase the accuracy of estimates, methods for combining estimates from different statistical models can also be used. Statistical methods used in the nowcasting context do not necessarily replicate those used in the regular compilation process, they mainly aim to produce the most reliable results for the reference variable, that is the best estimate of the evolution of the reference variable. Usually, they are applied to

the reference variable and their outputs are either growth rates on the previous period or on the corresponding period of the previous year.

When nowcasts are published, past values of the reference variable are usually not revised in order to avoid confusion among users. Such confusion could be generated by the differences between the nowcasted values and those coming from the regular compilation process, both on the information set and on the calculation methods. Nowcasts can be published either as part of the regular time series for the reference variable, or separately. In any case, nowcasts should not imply any revision of past values which should be revised only in the context of the regular production process. Nowcasting exercises should be carefully assessed, possibly in real-time, before taking any decision on their dissemination. Particular attention should be paid to the absence of a systematic bias between the nowcasted value and the first regular estimate of the reference variable and to their level of confidence. When possible, they should be published together with an assessment of the uncertainty implicit in their computation; a comparison with the first regular estimate could also be considered. Particularly attractive from a purely user point of view could be the publication of density nowcasts which provide users with an information about the expected value but also the degree of uncertainty (see chapter 14 of this handbook).

2.4.3 Flash estimates

Flash estimates aim to provide an early picture of the situation in a particular sector or in the whole economy based on one or several reference variables, possibly related by accounting and/or aggregation constraints.

Flash estimates will be computed after the end of the reference period, based on an incomplete set of information, at the moment when the amount of available information is considered sufficient to ensure a high degree of compatibility with the first regular computation. As in the case of nowcasting, the set of information can be incomplete in geographical, sectoral or temporal terms. However, in the case of flash estimates, the amount of partial information on the reference variable, or variables, should be larger than in the nowcasting case. In case the use of qualitative information would be necessary, it should be kept limited and mainly addressed to fill temporal gaps of some variables.

The information available when computing flash estimates should allow the possibility to replicate, even if in simplified way, the same compilation process used in the regular statistical process. Statistical methods used in compiling flash estimates should also be in line with the regular production process. In this context, the statistical methods can vary from simple aggregation techniques to regression methods, to temporal disaggregation and benchmarking ones, to grossing up techniques applied to small or incomplete sampling scheme, etc.

In the case of flash estimates, the use of combining alternative estimates techniques is not considered, unless it is also used in the regular compilation process. Flash estimates will usually lead to the estimation of the level of the reference variable, or variables; dissemination could however focus more on the period on period and year on year growth rates.

Since flash estimates will tend to approximate the same strategy used for the regular calculation, in principle they may produce a complete time series, with revised values for the previous periods. The opportunity of disseminating such revised figures has to be carefully evaluated case by case, assessing if disseminating past revised values or the latest value of the reference variable or variables only. In order to avoid any kind of confusion among users and to privilege the estimates obtained within the regular production process, it might be decided to publish the flash estimation of an additional observation of the target variable without any change of previous values.

Flash estimates should be assessed over a sufficiently long time interval, possible within a real-time exercise in order to analyse their behaviour with respect to the first regular calculation. They could also be benchmarked with the nowcasting of the reference variable and with the result of extrapolation methods in order to compare the relative performances of alternative rapid estimates strategies. Finally, as in the case of nowcasting, particular attention has to be paid to the absences of a systematic bias between the flash estimates and the

first regular estimates of the reference variable. Since deriving density estimates of flash estimates could be particularly complex due to the fact that they try as much as possible to mimic the regular production process, alternative ways to communicate their reliability have to be identified. In particular, flash estimates could be accompanied by the publication of their root mean square forecast error (RMSFE) or by the absolute mean revision with respect to the first regular estimates or, whenever possible, by a confidence interval.

2.4.4 Coincident and leading indicators

The approaches discussed in the previous sections aim to estimate the latest missing value for a given time series, which represents the main variable of interest. The methods presented in this section are based on a completely different philosophy. Those methods aim to compile a new time series (indicator) able to reproduce the same movements of the reference variable with a higher timeliness, instead of completing the reference variable with estimates for the latest points. Indicators can be classified into three main categories: coincident indicators, leading indicators, and lagging indicators.

Coincident indicators reproduce the movements of the reference variable without temporal shift, so that the movements at time t observed in the indicators correspond to those ones of the reference variable at the same time. Leading indicators are able to anticipate the movements of the target variable of a fixed number of periods. Lagging indicators reproduce the movements of the reference variable with some time-lags. Obviously, in the context of this handbook only the first two kinds of indicators are relevant.

Coincident indicators can be computed during the reference period or right after its end. Leading indicators are usually computed before the start of the reference period. Usually, coincident indicators produce one step ahead estimates where for leading indicators it is possible to produce multi-step ahead estimates, depending on the characteristics of the data used. Data involved in the construction of coincident and leading indicators can stem from different areas like survey data, financial data, real sector data, etc. They can also include variables from other economies than the reference one, or measuring relationships among economies. In principle, the amount of data used for the construction of coincident and leading indicators can be very large, up to thousands of series, which is another significant difference with the methods presented in the previous sections.

Estimation strategies for coincident and leading indicators are usually based on two-steps procedures where the two steps serve very different purposes, even if they can be simultaneously performed in some cases. The first step is devoted to the reduction of the variables space by the classification of variables according to their leading properties and the consequent variable selection. In this step data analysis techniques such as static and dynamic factor analysis and principal components, discriminant analysis, etc. can be used. This will lead to the identification of some latent variables (e.g. factors, principal components) to be used in the second step. Alternatively, approaches such as the general-to-specific modeling and the clustering analysis techniques can be used to select those variables with the highest explanation power. In the second step, a variety of methods from simple aggregation ones with fixed weights to regression based, bridge equations, VARs, linear and non-linear time series models, etc. can be used to estimate the coincident and leading indicators.

Coincident and leading indicators can supply a wider range of information with respect to the methods presented in previous sections. The previous methods were mainly oriented to the estimation of period on period or year on year growth rates and, in a more limited way, of levels of the target variable. Coincident and leading indicators can be targeted, thanks to the use of particular transformations applied to the components series, to the delivery of information covering the simple growth of the reference variable (period on period and year on year growth rates), its trend, cycle or growth cycle features, as well as the occurrence of turning points.

2.5 Rapid estimates vs. higher frequency estimates

For some key variables, traditionally available at quarterly or even annual frequency, an indirect way to increase their timeliness might be represented by the construction of new versions of such indicators available at higher frequencies. For example, for the GDP and the employment, which are usually available quarterly, the production of monthly related estimates could result on timelier low frequency estimates.

Of course, the construction of higher frequency indicators does not necessarily imply an improvement in the timeliness of the low frequency corresponding ones; however, from the user's point of view the availability of more frequent observations can be considered as a timelier monitoring of the economic situation.

Despite the fact that the practice of constructing higher frequency indicators is not yet very much accepted and used within statistical authorities, several interesting studies in this field have been produced since the 90's either by official statisticians or by academics and researchers. The focus of those studies has been mainly on the construction of a monthly GDP indicator (see Salazar et al. (1994), Barcellan et al. (1996), Frale et al. (2010) and Frale et al. (2011)), while only a few papers dealt with the construction of a monthly employment indicator (see Moauro (2014)).

When trying to extend the same techniques to higher frequencies, the level of complexity sharply increases; however, the amount of daily and weekly information and the availability of new data sources, commonly classified as big data, is constantly increasing, so that the possibility of producing in the near future very high frequency macroeconomic indicators, such as inflation and retail trade turnover (i.e. at weekly frequency), is becoming more and more concrete. In this line, Aprigliano et al. (2017) have tried to construct a daily indicator aiming to approximate the quarterly GDP growth rate for the current quarter. Chapter 9 of this handbook presents a survey of econometric models for mixed-frequency data.

A variety of statistical and econometric techniques can be applied to derive higher frequency indicators, from the more traditional ones, such as temporal disaggregation, to those based on structural models with missing observations (see chapter 8, and to mixed-frequency models (see chapter 9). Higher frequency indicators should be regularly benchmarked to their corresponding low frequency versions in order to ensure a full consistency across frequencies.

2.6 Rapid estimates and user needs

Statistical variables should serve specific purposes; they should reflect a need to measure or monitor a specific phenomenon of interest for the user, whether a policy maker, an econometrician, a student or any other kind of user, when such information is needed. Different estimates and indicators presented in the previous sections can be analysed in the wider context of users' needs.

Policy makers and in particular those deciding about macro-economic policies, need to have the freshest and most reliable information on the current economic situation; they are interested to know what are the variables' values in the specific moment they have to take a decision. They could also be interested in knowing how much the variable they are looking at is in the same phase of the business cycle or of other variables relevant in their policy context. In this respect, extrapolations will be usually considered with more carefulness, due to their typical univariate nature and the consequent inability to detect current turning points or breaks.

Flash estimates will constitute an useful instrument for a timely cross-country comparison but they do not necessarily fulfil the needs of policy and decision makers for real-time information.

Nowcasting will be a powerful instrument for real-time monetary and economic policy design, but it does not necessarily allow a reliable cross-country comparison due to the differences in the methodology used. The same apply to coincident and leading indicators which have the comparative advantage, with respect to nowcasting, of providing information for the current period or the next ones.

Users should be informed about the degree of reliability of all those kind of estimates and of the trade-off between timeliness and reliability. In order to fulfil users' needs, statistical authorities should be able to produce the complete variety of estimates covered in this chapter; however, this is rarely the case since if flash estimates are usually compiled by statistical offices, nowcasts and leading and coincident indicators are much less in the hands of official statisticians.

2.7 Conclusions

In this chapter an analysis of elements characterising different estimates in the context of official statistics has been presented. Starting from the definition of the elements necessary to identify in a unique way the concept of rapid estimates, a reflection on the methods used, the time of the estimation, the information used and the adherence to the official statistical production process have been introduced. The combination of those elements gives rise to a set of different estimates.

As a matter of fact, no standard terminology is widely accepted and a large variety of terms can be found in the literature. Terms as "early", "rapid", "flash", "first" estimate refer to similar concepts not distinguishable for the user. To adopt an agreed terminology would clearly foster common understanding and transparency. The way forward could be long before arriving to a wide acceptance of the proposed classification of estimates; however, not a long time ago the term nowcasting was almost unknown, while it has now become a commonly used one.

The chapter also includes a brief overview of an alternative way to foster timeliness based on the construction of higher frequency indicators. This provides the reader with the flavour of the wide variety of products, and associated statistical techniques, which can be classified under the umbrella of rapid estimates. Rapid estimates are then framed in the general context of users' needs. Finally, despite the fact that ideally all kind of rapid estimates above presented should be produced by statistical authorities, this handbook will concentrate on rapid estimates covering the recent past or the present and will not deal with feature estimates such as those derived from leading indicators.

Annex

2.A Glossary of statistical terms associated to the estimation of key economic indicators

Coincident indicators: Economic and financial market indicators which tend to move in step with general economic trends such as gross domestic product, employment levels, retail sales, and/or financial market trends such as interest rates and stock market prices.
Synonyms: concurrent indicators
Source: BusinessDictionary.com

Estimate: In the strict sense, the particular value yielded by an estimator in a given set of circumstances.
Source: The OECD Glossary of Statistical Terms

Estimator: A rule or method of estimating a parameter of a population.
Source: The OECD Glossary of Statistical Terms

Extrapolation: Statistical technique of inferring unknown from the known. It attempts to predict future data by relying on historical data, such as estimating the size of a population a few years from now on the basis of current population size and its rate of growth. Extrapolation may be valid where the present circumstances do not indicate any interruption in the long-established past trends. However, a straight line extrapolation (where a short-term trend is believed to continue far in into future) is fraught with risk because some unforeseeable factors almost always intervene.
Source: BusinessDictionary.com

Flash estimate: An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.
Synonyms: advance estimate (BEA); preliminary estimate (ONS), rapid estimate (Ottawa conference)
Source: Eurostat Handbook on Quarterly National Accounts

Forecasting: and prediction: often used synonymously in the customary sense of assessing the magnitude which a quantity will assume at some future point of time: as distinct from “estimation” which attempts to assess the magnitude of an already existent quantity. For example, the final yield of a crop is “forecast” during the growing period but “estimated” at harvest.
Source: The OECD Glossary of Statistical Terms

Hard data: Reliable and methodologically sound data from official or organizational statistics that is comparable and roughly independent from the way it was measured.
Synonyms: factual, quantitative data
Source: Eurostat

Interpolation: Estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information. Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. Time series analysis and regression analysis are the two statistical techniques employing the concept of interpolation.
Source: BusinessDictionary.com

Nowcasting: A very early estimate produced for an economic variable of interest over the most recent reference period calculated on the basis of incomplete data using a statistical or econometric model different from the one used for regular estimates. Soft data should not play a predominant role in nowcasting models. Nowcasts may be produced during the very same reference period for which the data is produced.
Source: Eurostat

Soft data: Other than hard data, generally less reliable.
Synonyms: qualitative data
Source: Eurostat

Regular estimate: Estimate produced and published for an economic variable of interest over all reference periods normally produced and published using a statistical or econometric model based largely on hard and complete data.
Source: Eurostat

2.B A proposal for a common glossary of rapid estimates[2]

2.B.1 Introduction

Statistics are used as daily reference and source of information to interpret how economic and social phenomena are evolving. Consequently, a wide range of timely and reliable data is required to monitor the evolution of the business cycle in order to assess its impact on all parts of society. Official statistical authorities usually deliver high quality data, however with some delay. The importance of the timeliness/reliability trade-off raises many technical and practical issues. Here we adopt a more general approach focusing on the conceptual difficulties related to this trade-off and the production of "rapid estimates". Because the content of this notion will be at the core of this annex, let us for the moment use the term "rapid estimate" as a generic concept including several estimates which will be precisely defined in this glossary.[3] We start with some definitions:

Definition 1. **Rapid estimates (or early estimates)**: General expression denoting a timely numerical evaluation (i.e. an estimate) of an economic variable of interest, using the real-time data flow, such that an early picture of the present, the recent past or the near future can be formed.

Definition 2. The **real-time data flow** at a specific point of time is the whole information set available to statistics producers (official statistician, researchers,...). It includes soft, hard, financial, unconventional data and it usually provides an incomplete data coverage of the reference period at the specific point of time. The real-time data flow is by construction unbalanced: ragged or jagged hedge data.

Note that the characteristic of being unbalanced is a consequence of the use of data with different frequencies; when looking at quarterly data, some others monthly indicators could be available before the end of the quarter we are referring to.

A clear communication on the definition and the quality of different types of statistics is needed to guarantee the credibility expected from official statistics. For example, rapid estimates may include nowcastings, flash estimates, advance estimates, preliminary estimates, first and subsequent releases and, under some conditions, forecastings, or coincident and leading indicators.

To facilitate the discussion and the exchange of opinions among national statisticians and other users (central banks, academic researchers, etc.), a common operating terminology has to be adopted and used. For

[2] Authors: Roberto Barcellan, Alain Hecq, Gian Luigi Mazzi and Rosa Ruggeri Cannata.
[3] Note that in the literature the term "early estimates" can also be found. Although it is a synonymous of rapid estimate, we shall only use the latter in this document to avoid any confusion.

instance, although in the academic world, some of these concepts are referred to as “forecast”, for many official statistics producers (OSPs) the terminology of “rapid estimates” is often preferred.

While the term forecast is not precise enough, the latter terminologies, although intuitive, are also unfortunate. Indeed, how should the estimates be defined? They clearly cannot be called “slow”, “normal” or “late” estimates.

This annex is structured as follows: section 2.B.2 introduces some general concepts at the base of our framework; section 2.B.3 details the different axes and sub-branches which constitute our framework, while section four presents definitions of estimates on the base of the concepts introduced; finally section five concludes.

2.B.2 Background concepts

In the previous section we introduced some lack of an harmonized language and the need for developing a common terminology. In this section we further detail first such lack of harmonisation by showing concrete examples and propose then a methodology based on the axes/dimensions/layers which we consider necessary to define a clear framework and avoid the potential lack of consensus in the labelling of terms for rapid estimates (see Table 2.B.1).

Table 2.B.1: An example of terms used for the first three estimates of the GDP.

	1st release	2nd release	3rd release
Eurostat before 2011	flash estimate (+45)	1st regular rel. +65	2nd reg. rel. +105
Eurostat 2011 onwards	flash estimate (+45)	2nd estimate +65	3rd estimate (+95)
BEA	advance est.+30	preliminary estimate +60	final or 3rd estimate (+90)
ONS	preliminary +25	2nd estimate M2 (+55)	month 3 estimate (+85)

BEA=Bureau of Economic Analysis (US), ONS=Office for National Statistics (UK).

The introduction of axes (multifaceted dimensions) first aims to avoid, or at least minimize, confusion of terms leading to an almost unique identification of various types if rapid estimates. The focus is on the statistical terms associated to the evaluation of key economic indicators (e.g. Principal European Economic Indicators = PEEIs) and in particular on terms that are commonly in use but in most cases with different interpretations. In this framework, clear definitions are going to be provided for such terms as rapid estimates, flash estimates, nowcast, forecast, advance estimate, real-time data flow, parameter uncertainty, information set, reliability, etc. with the aim of making such definitions commonly accepted.

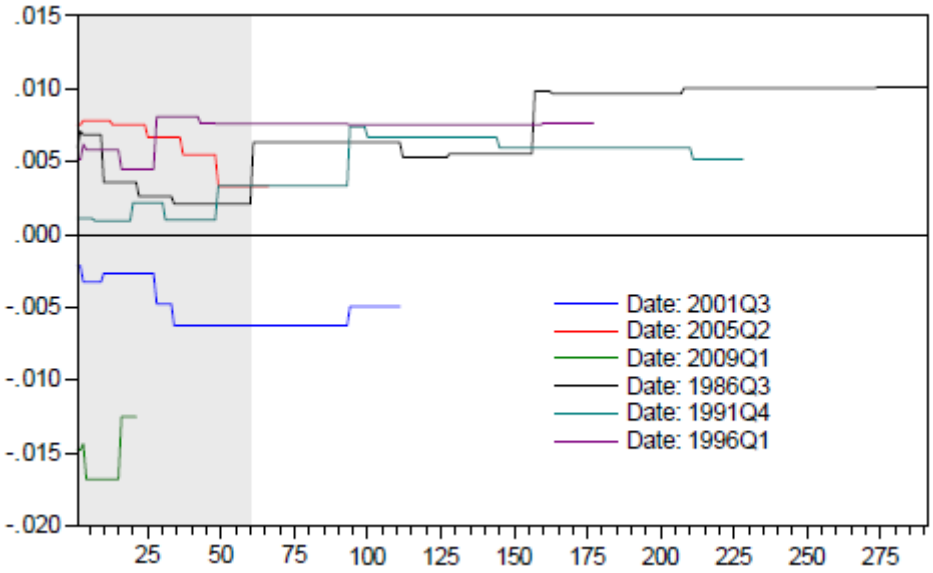
Timeliness versus reliability

We will start our analysis of concepts related to estimates by separating the timeliness concept from the reliability one. Moreover, in order to keep the task feasible, we need to restrict the time horizon for the comparison: it would not be significant to compare an estimate based on partial information with a final release published in 10 years from now.

As an example, the following graph shows the quarterly growth rate of real GDP (seasonally adjusted) in the US for several quarters (calendar dates) but evaluated for 300 monthly vintages (publication releases). Let us take for instance the calendar date 1986Q3. The evaluation made a few months after the end of the quarter (first few releases) is by chance similar to the one obtained after 5 years (end of gray box), although there exists a U-shape between these two dates.

However, both evaluations are very different to the final (last) release. This well know result emphasizes that a numerical evaluation of rapid estimates should be done a posteriori (e.g. reliability of the first versus the final release), however this cannot be the base for defining concepts.

Figure 2.B.1: Quarterly seasonally adjusted real GDP - US



Possible Axes/Layers

Based inter alia on previous work by Eurostat (see Barcellan et al. , 2009), we have identified the following potentially relevant axes/layers:

- Single producer of the official release versus the multiplicity of the producers
- Prerogative of a statistical authority versus a non Official Statistical Producer
- Timeliness and data availability. The relative position in time of the estimation with respect to the reference period
- Econometric versus statistical model (model uncertainty and/or parameter uncertainty)
- Information set used for the estimation
- Regular estimate versus non regular estimate
- The target variable to be assessed (economic indicators vs. proxy), statistics/proxy variables (e.g GDP/confidence consumer indicator)
- Out-of-sample versus in-sample evaluation
- Aggregated versus disaggregated forecast
- Frequency (periodicity) of the variable of interest (annual/quarterly/monthly)
- Flow/stock time series
- Seasonally adjusted versus non seasonally adjusted data
- Timing and frequency of the vintage (published or internal information)
- Transformed/untransformed data (e.g. first difference, growth rates or level)
- Revision/final release
- Hard/soft explanatory variables
- Partial coverage/full coverage

- Survey/census
- Univariate/multivariate model for forecasting the variable of interest.

In the next section we will analyse them to derive a simplified subset allowing the definition of different kind of rapid estimates.

2.B.3 A complete framework based on eight axes

In this section we propose a reduction of the number of axis to eight, classified according to:

- **Who** makes the evaluation (1 axis)
- **What** is evaluated (2 axes)
- **How** is the evaluation done (3 axes)
- **When** is the evaluation done (2 axes)

Let us develop these layers in more details.

Axis 1: The uniqueness of an official release versus the potential multiplicity of producers

This axis is about who produces the variable of interest. The question raised here is: Is the institution who produces the rapid estimate also the usual (official) supplier of the regular estimate? The producer of a rapid estimate may or may not be the same as the producer of regular releases for a given indicator. This first criteria is the simplest and the most obvious one (although not enough) for making a clear cut among some definitions of rapid estimates. It concerns who is the producer, namely the responsibility/liability of the institution which produces the statistics.

Institutions are seen in a broad sense, i.e. we look at Official Statistical Producers (OSP from now on); they can be National Statistical Offices (such as the Bureau of Economic Analysis (BEA, United States), the Office for National Statistics (ONS, United-Kingdom), the Institut National de Statistiques (INS, Belgium), in some cases a central bank or a planning office on behalf of them, international organizations (Eurostat, IMF, OECD, World Bank), a national ministry, local institutions, sectoral or company organizations, unions, and so on. Hence it is only when the institutions producing rapid estimates are also those producing regular ones, that we can call their estimates "Flash estimate", 1^{st} or 2^{nd} release.

This first condition is necessary but not sufficient to distinguish among several estimates. Non official statistical authorities, such as forecasting centers, academic researchers or experts can and do compute and provide forecasts or nowcasts; however, there are some other mixed cases: an institution could not be the official producer of a regular estimate, and provide forecasts or nowcasts. It could also be the case that the official producer of a regular estimate does produce a forecast using a different methodology from that used for the regular production. This shows that the concept expressed in this axis is not sufficient to distinguish among several estimates.

Moreover, we would like to notice that in order to foster clarity, it is important that when an institution publishes a figure from a third party, this figure is disseminated with a clear indication of the source and the same name that has been given by the original producer (e.g. a flash estimate from OECD published by a forecasting institute should be clearly identifiable).

Let now go in more details about what can be considered an official release. Imagine that in a country, data are collected by the national statistical office and, due to particular agreements, the National Bank computes figures on behalf of the National Statistical Office too. Those figures can be part of official releases. The situation is different from the case where a financial institution decides to produce a newsletter for his customers and predicts the movements in the most recent economic activity. This is also different from the case where a planning office is charged by a national government to give an estimation of the economic activity to discuss the national budget (outside the official figures). All these cases cannot be called official releases.

Axis 2. The target variable (What 1)

This axis is about 'What is evaluated'. Based on Axis 1, only data produced by an Official Statistical Producer (or on behalf of it) can be officially "released". When computing a rapid estimate, the target variable can be hard, soft, financial or unconventional.

The following definitions aim to clarify what is meant for hard, soft or financial data.

Definition 3. Hard (or factual) data (Eurostat) : reliable and methodological sound data from official or organizational statistics that is comparable and roughly independent from the way it was measured.

Definition 4. Soft data: data in the form of qualitative or quantitative information resulting from an approximation of an economic phenomenon by survey and polls.

Therefore, soft data are dependent on the way they are collected, for instance by survey.

Definition 5. Financial data: interest rates, stock prices, exchange rates, and similar data

Definition 6. Unconventional data: experimental data, data under development and particular useful indicators (i.e. business cycle turning points).

The distinction between hard and soft data is important when they are used in the information set (input series) to compute another series because they could be the instruments helping to construct different kinds of rapid estimates.

Axis 3. Revisions in the estimate (What 2)

Rapid estimates concern data that are forecasted or released with the idea that they are preliminary figures that will be likely revised after their first release. Financial and opinion surveys data are important economic indicators, that are quickly available and are not revised; they could however be forecasted and in this case the terminology we are introducing applies.

Definition 7. Revision (Eurostat): Revisions are broadly defined as any change in a value of a statistic released to the public. They can occur either when new observations become available and some past values are modified or when the current and/or some past values are modified

Data currently produced by OSPs are often subject to a recurrent revision process resulting in the publication of successive releases, which represent different quantitative evaluations of the same phenomenon under study. There are various reasons why the revision process is necessary. For example, revisions can be caused by improvements in the estimation techniques used by OSPs, enlargements of the information set, refinements in the system of basic statistics, changes in the accounting definitions and classifications, or by changes in the base year for constant prices evaluations.

After having clarified what kind of data we are talking about in the input process of an estimation procedure, we can now move to examine aspects about how an estimate is obtained.

Axis 4. The adherence to the regular production process (How 1)

When assessing this dimension we will look at the degree of adherence of the rapid estimates to the regular ones. The first concept we would like to introduce concerns the production process of a regular estimate.

Definition 8. Regular estimate (Eurostat): a regular estimate is an estimate produced and published for an economic variable of interest over all reference periods, usually produced using statistical techniques based largely on hard and as complete as possible data.

Rapid estimates can be based on a process fully or partially adherent to the production one, or substantially different. The degree of adherence or not to the regular production process is crucial in discriminating, for instance, between flash estimates and nowcasting.

For example, let us define the quarterly euro area GDP as the sum of the GDP from countries belonging to the euro area (for the sake of simplicity we suppose the sum is not weighted, this will not change our reasoning).

Let T be the end of a reference quarter, and consider an observer at time $T+\varepsilon$ ($\varepsilon > 0$ and small) with the data flow available in $T+\varepsilon$ (denoted Ω_T), who should produce an estimation of the variable of interest at the end of the quarter T, y_T. $T+\varepsilon$ simply means a few days (or weeks) after the end of the reference quarter before the end of the next quarter $T+1$.

Then, if for the quarter $T-1$ we have a simple aggregation scheme such that

$$y_{T-1} = y_{1T-1} + y_{2T-1} + ... + y_{nT-1}$$

The regular production process implies that next quarter will use the same scheme

$$y_T = y_{1T} + y_{2T} + ... + y_{nT}$$

and not any other function

$$\hat{y}_T = f(\Omega_T)$$

Nevertheless, since Ω_T will be less complete than Ω_{T-1}, the gaps should be filled in (e.g. estimation of missing countries at time T) before applying the aggregation scheme; this can be considered as an example of a rapid estimate fully adherent to the regular production process. By contrast, if at $T+\varepsilon$ we aggregate only the available countries at time *T* and we complement this partial aggregation process with a grossing up procedure, then the obtained rapid estimate is based on a process partially adherent to the production one. When we wish to emphasize the use of the same information set as in the regular production, we will indicate it by $reg(\Omega_T$.

Axis 5. Information set (How 2)

In order to complete our analysis of how a figure is produced, we need to better clarify the role of the information set and of its degree of completeness. Continuing to develop our previous example, when the information set is incomplete, we obviously allow for some kind of estimation, i.e.

$$\hat{y}_T = \hat{y}_{1T} + \hat{y}_{2T} + ... + \hat{y}_{nT}$$

under some specific conditions of uncertainty. However to obtain the elements of $\hat{y}_{1T}, \hat{y}_{2T}, ..., \hat{y}_{nT}$ evaluated at $T+\varepsilon$, the information set must include the real-time data flow in T (i.e. Ω_T and not only Ω_{T-1}).

Note that this requires to define a coverage rate for the aggregate value in terms of the n components. By coverage we mean the extent to which the total aggregate is covered by the sampling. An example of incomplete observation set is the case of two monthly values available when estimating a quarterly aggregate. It is also useful to notice how the information set used for producing rapid estimates can also include variables which are not among those used in the regular production process; for example the presence of opinion surveys data in information sets used for rapid estimates is justified by the lack of hard data. When hard data progressively become available the presence of soft data should decline accordingly.

Axis 6. Model/versus parameter uncertainty (How 3)

This is the third concept necessary to identify 'how is a figure produced'. As a first idea, we could be tempted by introducing statistical and econometric models; however, the distinction between an econometric and a sta-

tistical model is quite controversial since there are different interpretations from econometricians and statisticians. Let us first clarify the notion of uncertainty. We will talk about model uncertainty and parameter uncertainty instead of econometric and statistical models. Looking at model uncertainty, it is well known that model-based estimates and forecasts are clearly conditional on the chosen model. However the choice of an appropriate model is itself subject to uncertainty (ARMA, Kalman filter, regression model and with which explanatory variables,..). Sometimes the model specification relies on strong a-priories while in other cases is the result of a statistical model selection procedure.

Even if we assume that the model is known, there is still a source of uncertainty about the estimation of the parameters due to measurement errors and/or estimation methods used. The second source of uncertainty, i.e. statistical uncertainty, is due to the unknown parameter estimates that we obtain from a finite sample. Let us also call the true underlying phenomenon the data generating process, namely the **DGP**. We define "simple arithmetic rules" smoothing techniques based on basic rules such as ratios, moving averages, etc.. Basic rules are not necessary trivial, as it is the case of X11 filters. The following table summarizes the different sources of uncertainty characterizing various methods, allowing by consequence a clear distinction between statistical and econometric ones. Looking at the table, cells (1) and (2)(i) refer to statistical models while cells (2)(ii). and (3) refer to econometric models.

Table 2.B.2: Different forms of uncertainty

	DGP	Parameters		Estimation technique
(1)	known	known		interpolation, extrapolation
(2)	known	estimated	(i)	simple arithmetic rules, ratios
			(ii)	statistical/economic techniques
(3)	unknown	estimated		

Axis 7. An appropriate release time (When 1)

We focus now on the timing of the release of a rapid estimate. Rapid estimates could be compiled before the reference period, during the reference period, around the end of the reference period or later. Indeed not all estimates produced after the reference period can be qualified as rapid estimates. For example, if a subcomponent of the national accounts is first released two months after the reference quarter, it cannot be considered as a flash estimate but as a regular estimate.

Obviously there is not a clear borderline between rapid and regular estimates, it will depend on the frequency of the observation and also on the characteristics of each national statistical system. We assume here, based on practices observed in several countries, that rapid estimates should be released, for infra-annual data, within the first half of the following period. A rapid estimate for y_T, should then be released not later than at $T + \frac{1}{2}\Delta T$, where ΔT denotes one unit of the observation frequency (i.e. one month or one quarter). This means a maximum of 15 days for monthly data, and of 45 days for quarterly ones. This choice is somewhat arbitrary and could be adjusted according to specific cases. Indeed, it seems obvious that data of a census on households carried out every 10 years is not going to be considered rapid estimates if published before 5 years after the end of the census. For annual data, it also depends on whether we have stock or flow data. We expect a rapid evaluation of the annual GDP because it is based on the aggregation of quarters. We do not expect however such an early evaluation for migration data or environmental indicators.

Axis 8: Stock and flow data, Reporting and reference period (When 2)

We investigate now the difference between the reference and the report period. For flow data, they obviously coincide so that is it impossible to have an estimate with a complete coverage before the end of the period T. For example, to obtain the number of cars registered in December 2015 we need to wait until the end of the month (assuming figures are instantaneously available). For stock data, reference and reporting period do not

necessary coincide, mainly because data can be recorded in a given day or week within the reference period. In this situation an estimate with full coverage could be produced even before the end of the reference period. We prefer then to refer to reference period and reporting, or collecting, period.

The following table gives a synthetic presentation of the eight axes introduced with their modalities, i.e. the values which should be considered as alternatives for those axes.

Table 2.B.3: Axis and modalities

	Axis	Modalities
Who?	1. The uniqueness of an official release versus the potential multiplicity of producers	Statistical offices or members of the statistical system (OSPs) Other governmental institutions Private institutions
What?	2. The target variable	Hard data Soft data Financial data Unconventional data
	3. Revisions in the estimate	Data subsequently revised Not revised data
How?	4. The adherence to the regular production process	Fully adherent to the regular production process Partially adherent to the regular production process Different from the regular production process
	5. The Information set	Availability of the full information set for the period under estimation Incomplete observation set for the period under estimation No available information for the period under estimation
	6. Model versus parameter uncertainty	The model is known or chosen, parameters are known The model is known or chosen, parameters are estimated Both the model and the parameters are chosen to evaluate the variable of interest
When?	7. An appropriate release time	Estimates produced before the reference period Estimates produced during the reference period Estimates produced after the end of the reference period, but not later than $T + \frac{1}{2}\Delta T$ Estimates produced after $T + \frac{1}{2}\Delta T$
	8. Reporting and reference period	The reporting period and the reference period coincide The reporting period takes place within the reference period

2.B.4 Main definitions

In this section we present as examples some definitions of special kinds of rapid estimates as derived from the axes that were introduced before.

Flash estimates

Several definitions of flash estimates can be easily found and are reported below. We will then present the glossary definitions of flash estimates.

Definition 9. A flash estimate is an early estimate produced and published as soon as possible after the end of the reference period, using a more incomplete set of information than the set used for traditional estimates – (Eurostat, Ottawa seminar)

Definition 10. GDP flash estimate: the earliest picture of the economy according to national accounts concepts, which is produced and published as soon as possible after the end of the quarter using a more incomplete set of information than that used for traditional quarterly accounts.

Definition 11. A flash estimate is an early estimate for an economic variable of interest over the most recent reference period and is normally calculated on the basis of a statistical or econometric model. The flash estimate should have a release date appreciably earlier than the first release date of the actual data for that variable (available on Flash ESSnet – Eurostat).[4]

Definition 12. Flash estimate (Eurostat) : An early estimate produced and published for an economic variable of interest over the most recent reference period. The flash estimate is normally calculated on the basis of incomplete data, however produced using the same statistical or econometric model as for regular estimates. Flash estimate models should incorporate hard data as much as possible.

Definition 13. GDP Flash Estimates are produced using a more incomplete set of information to give the earliest picture of the economy. It is considered as more timely as compared to GDP announcements. However, some economists are skeptical of the flash statistics as they may trigger an unwarranted reaction in the stock and bond markets. (Wikipedia)

The following definition is based on all the dimensions above introduced, and clearly identifies the concept of flash estimate in a unambiguous way. The definition provided below includes explicit specific references to the eight axes.

Definition 14. Flash estimate (according to this Glossary): A flash estimate for the most recent reference period T is the first official release of generally a hard economic indicator (Axis 2) computed **before** $T + 1/2\Delta T$ (Axis 7) and produced by the statistical institution in charge of the regular estimate (Axis 1), using an incomplete set of information (Axis 5) exploiting as much as possible all available hard data (Axis 2). A flash estimate is produced by using a methodology as adherent as possible to the one used for the regular estimates (Axis 4). Statistical techniques (Axis 6) can help in adjusting the temporary (Axis 3) incomplete observations. While hard data (Axis 2), if they are observed during the reference period (Axis 5 and 8) should be favored as much as possible, soft data (Axis 2) might be also used.

From this setting other definitions can be derived, in line with the introduced structure.

Definition 15. First release: a first release for the most recent reference period T is the first official release of generally a hard economic indicator (Axis 2) computed **after** $T + 1/2\Delta T$ (Axis 7) and produced by the

[4] See http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Flash_estimate.

statistical institution in charge of the regular estimates (Axis 4). The first release is produced according to the statistical methodology adopted for the construction of regular releases. Statistical techniques (Axis 6) can help in adjusting the temporary (Axis 3) incomplete observations, if necessary. Hard data observed during the reference period (Axis 5 and 8) should be the core of the estimation.

Note that the difference between a flash estimate and a first release is the time at which the series is reported and the prominent role of hard data.

Definition 16. Second release: The second release is the second official release of an economic variable of interest for the most recent reference period. It concerns the release of generally hard data that are subject to possible further revisions but produced according to the statistical methodology adopted for the construction of regular estimates. This release updates information first published in the flash or first estimate by using a higher coverage of the sub-aggregates. It includes more detailed output data than the flash or first estimate. The role of hard data must be dominant.

In a similar way, successive releases can be easily defined, including maybe some precision about the coverage, the content, etc.

Definition 17. Third release (final estimate): The third release (sometime final estimate) is the third official release of an economic variable of interest for the most recent reference period. It concerns the release of generally hard data produced according to the statistical methodology adopted for the production of regular estimates. This release updates information published in the second estimate by using in most cases a full coverage of the sub-aggregates. It includes detailed output. The role of hard data is compulsory. In most case this third release can also be called the final estimate which doesn't mean that the figure will not further be revised when more accurate information is made available.

Advance Estimate

Using the same framework introduced in this glossary, it is interesting to define an advance estimate as follows. The notion of advance estimate might be preferred to the one of preliminary estimate.

Definition 18. Advance Estimate (according to this Glossary): For an institution in charge of the regular estimate (Axis1), an Advance Estimate for the most recent reference period T is an estimate of generally a hard economic indicator (Axis 2) computed **during** the reference period T (Axis 7). Although it is calculated on the basis of an incomplete information set it is produced using the same methodology adopted for the production of regular estimates (Axis 4). Statistical and econometric techniques (Axis 6) are used to fill in the temporary (Axis 5 and 8) incomplete observations. Hard, soft, unconventional and financial data are used if they are observed during the reference period T.

In some cases, a producer could be able to partially evaluate the current state of a series for the reference period T using some partial information available in T and considering the adherence principle to the regular estimate, but with the support of modeling techniques to fill remaining gaps. In this case, we would consider such estimates as an advance estimate. This is different from the nowcasting issue that we analyze next.

Nowcasting

Banbura et al. (2010) define nowcasting as forecasting the present, the near future or the recent past. This interpretation could be considered for an external observer/forecaster but not for an institution in charge of the production of the regular estimates (Axis 1). And this makes nowcasting very different from flash and advance estimates. We define nowcasting as follows:

Definition 19. A **Nowcasting (according to this Glossary)** is a rapid estimate produced by a statistical authority or an institution outside a statistical system during the current reference period T (or very close to

its end "the present") for a hard economic variable of interest observed for the same reference period T. A nowcasting makes use of all available information becoming available between $T-1$ and T until the estimation time. Statistical or econometric models, different from the ones used for the regular production process are considered using either hard, soft, unconventional and/or financial data.

A nowcasting is then a particular case of ex-ante forecasts of the present using the real time data flow available between $T-1$ and T. When the information set includes data up to $T-1$ only, we will have an ex-ante forecast. Nowcasting differ from advance estimates because the latter is obtained using the same scheme (adherence principle) as the one used for the regular production process.

It is worth to notice that the nowcasting term is used in many different ways in scientific papers. The definition provided here is functional to the production of rapid estimates by statistical institutions. In the rest of the handbook the term nowcasting is used in a different way, reflecting the views of chapters' authors.

A comparison by examples

The following tables summarize some of the introduced definitions focusing on the moment in time when the estimate is produced. Let us first consider a case where an OSP evaluates the GDP for time T in $T-1$, T or $T+1$ with $\Omega_T^* > \Omega_T$. The classification of the type of estimate will depend upon a combination of factors reflecting the axis introduced.

Table 2.B.4: GDP rapid estimates prodcued by an official statistical producer

T-1	evaluation in T		evaluation T+1	
ex-ante forec.	nowcast	if f(Ω_T)	Flash if reg(Ω_T^*)	first rel.
	advance est.	if reg(Ω_T)		
	forecast	if f(Ω_{T-1})		

Compare this to the case where a research center forecasts the GDP for T in $T-1$, T or $T+1$:

Table 2.B.5: Non-official GDP rapid estimates

T-1	T		T+1
ex-ante forecast	nowcast	if f(Ω_T)	ex-post forecast
	forecast	if f(Ω_{T-1})	

Again, we will reach different results on the base of the information set and the production process.

2.B.5 Conclusions

In this section we have introduced a glossary on rapid estimates by introducing a framework for the classification of estimates based on eight axis with related modalities. This framework serves the aim of increasing clarity in the terminology linked to rapid estimates. Although it is very challenging to come to unambiguous and commonly accepted definitions, we have clarified all dimensions necessary to this task. The set of axes could appear cumbersome at a first look, but the examples show how they are really needed to uniquely classify the wide range of existing estimates.

Bibliography

Aprigliano V., Foroni C., Marcellino M., Mazzi G.L., Venditti F. (2017). A daily indicator of economic growth for the euro area, International Journal of Computational Economics and Econometrics (IJCEE), 7(1-2), pp. 43-63.

Bańbura, M., D. Giannone, and L. Reichlin (2011): "Nowcasting" in Oxford Handbook on Economic Forecasting, ed. by M. P. Clements, and D. F. Hendry, pp. 63 - 90. Oxford University Press.

Barcellan R., Bruno G., Mazzi G.L.(1996). Quarterly National Accounts in EU Countries: Situation and Perspectives, notes of TES course on Quarterly National Accounts.

Barcellan R., Mazzi G.L., Montana G. and Ruggeri Cannata R. (2009). Glossary of statistical terms associated to the estimation of key economic indicators, International Seminar on Early Warning and Business Cycle Indicators, Scheveningen, The Netherlands. December 2009.

Frale C., Marcellino M., Mazzi G.L. and Proietti T. (2010). Survey data as coincident or leading indicators, Journal of Forecasting, 29(1-2), 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, 439-470.

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.

Moauro F. (2014). Monthly Employment Indicators of the Euro Area and Larger Member States: Real-Time Analysis of Indirect Estimates, Journal of Forecasting, John Wiley & Sons, Ltd., vol. 33(5), pages 339-349, 08.

Salazar E., Smith R., Weale M., Wright S. (1994). Indicators of monthly National Accounts, presented at Eurostat - INSEE Workshop on Quarterly National Accounts, Paris 5-6 December 1994.

3 Forecasting and Nowcasting Macroeconomic Variables: A Methodological Overview

Authors

Jennifer Castle
Magdalen College Oxford
jennifer.castle@magd.ox.ac.uk

David Hendry
Nuffield College Oxford
david.hendry@economics.ox.ac.uk

Oleg Kitov
St Hilda's College Oxford
oleg.kitov@economics.ox.ac.uk

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

3.1 Introduction

Nowcasting is 'forecasting' the current or recent aggregate state of an economy. It can be undertaken either at the initial stage of improving 'flash' estimates within a National Statistical Agency, which would otherwise just be based on inferences from the intermittent arrival of highly disaggregate data measures, or by an outside organization using higher-frequency available information such as surveys to improve available aggregates, or both. Here we focus mainly on the former, and explain the use of recent developments in both modeling and forecasting to improve the accuracy and timeliness of initial data releases. Approaches based on ascertaining the state of an economy from a wide range of information sources, so 'high-frequency estimates' of relevant aggregates are always available, are discussed by Angelini et al. (2011) and Giannone et al. (2008), with a survey in Bánbura et al. (2011). Some of the issues and methods discussed here apply to both approaches. We refer to any procedure that uses additional information when producing contemporaneous aggregate data, beyond just cumulating observed disaggregates, as nowcasting.

There are obvious practical reasons why up-to-date information is needed on economic aggregates, especially those which play a key role in economic policy decisions. There are four difficulties confronting the production of timely and accurate low-frequency aggregates, necessitating nowcasting, as we now explain. A fifth problem (denoted the 'breaks problem' below) can adversely affect the accuracy of all forms of nowcasts. There are two other difficulties facing the production of 'continuous high-frequency estimates' of aggregates. We address these seven issues in turn.

First, not all disaggregated contemporaneous data are available at the time their information is required to construct a relevant aggregate–the **'missing data problem'**. Aggregates cannot be constructed just by adding up the relevant observed disaggregates at the end of the relevant period. For preliminary estimates of UK GDP, for example, the information is only available up to 24 days after the end of the quarter for which a flash estimate is required: see Reed (2000) and Reed (2002). This is the central basis for 'nowcasting' at any Statistical Agency, essentially forecasting the missing disaggregates.[1]

Secondly, a **'latency problem'** is ever present, in that some variables have to be estimated or 'in-filled' even long after the end of the period just noted. A 'nowcast' of those disaggregates could lead to a more timely and accurate aggregate measure.

Thirdly, different components of the data entering the aggregates for which nowcasts are required are unavailable in different periods, and so are missing on a non-systematic basis: see e.g., Clements and Hendry (2003). Consequently, a consistent subset of information is rarely available, inducing a **'changing database problem'**, which is bound to affect the design of any system for nowcasting.

Fourthly, many disaggregate economic time series are themselves only preliminary estimates, which are subject to potentially substantial revisions in due course as more information accrues, so are not necessarily a reliable and accurate guide to current conditions. Faust et al. (2007) report that revisions to GDP announcements are quite large in all G7 countries, and are fairly predictable in the UK, mainly because of reversion to the mean, which they interpret as due to removing measurement 'noise'. Thus, **'measurement error problems'** are also pervasive, and consequently, even the 'observed data' cannot be taken at face value, and may be no more accurate than nowcast values.

Fifthly, a **'breaks problem'** from unanticipated location shifts (changes in the previous means of the variables under analysis) always threatens to disrupt the accuracy of nowcasts, whether based on 'in-filling' for missing disaggregate data, or the production of 'continuous high-frequency measures' of aggregates. Careful attention is required to avoid shifts contaminating flash estimates. In section 3.6, we propose that 1-period ahead forecasts of the disaggregates should always be produced, enabling immediate evaluation against incoming measured outcomes, to act as an 'early warning signal' when measured series and forecasts depart radically. Comparisons would facilitate rapid action to be taken following any deterioration in the performance of a

[1] A separate 'missing data problem' is that the existence of some products, companies, etc., may be unknown to the National Statistical Agency at the time, and only incorporated into the data base much later: we do not address that issue here. This is also an extreme version of the 'measurement error problem' discussed in section 3.6.8.

nowcasting model or method, or warn of measurement problems in the 'actual' series. A cross check could be based on robust forecasts as discussed in section 3.4.3, and this applies to both nowcasting approaches.

Next, to produce 'continuous high-frequency updates' of aggregate measures requires handling a potentially vast amount of data, which we denote the **'problem of more variables, N, than observations, T'**. In fact, this difficulty can also apply to methods aimed at improving aggregate flash estimates directly. Possible solutions include data reduction (using automatic model selection), data combination (as in principal components or factor methods), and model combination (as in model averaging).

Last, some of the high-frequency data to be used to estimate the low-frequency aggregate are not released in a systematic or synchronous way, and may have different publication lags with each release. Like the 'changing database problem', this **'non-synchronous release problem'** necessitates a flexible software system.

The interaction of these seven problems poses major uncertainties for nowcasters. At first sight, the difficulties are similar to those facing forecasters, and indeed we will draw on the literature analyzing forecasting during breaks. However, a key difference is the presence of contemporaneous observations—which logically could never occur in a forecasting context—and there are many possible sources of such evidence, including rapidly and frequently observed data on variables such as retail sales, house prices and financial markets; correlates like road traffic and air passenger numbers or energy consumption; surveys; and more recent innovations, including Google Trends described by Choi and Varian (2012), mobile phone data, and prediction markets: see Wolfers and Zitzewitz (2004).

Castle, Fawcett, and Hendry (2009) and Castle and Hendry (2010) propose a framework for nowcasting that addresses the issues of shortening the 'latency', in-filling for 'missing data' despite ongoing shifts, reducing 'measurement errors', while being sufficiently flexible to handle 'changing databases' and 'non-synchronous releases' yet tackling the 'breaks problem'. Their approach exploits many sources of contemporaneous data in an automatic model selection algorithm designed to handle $N > T$, as well as tackle unanticipated location shifts both directly and by using robust forecasting devices. We explain the various ingredients of their approach, and illustrate by an application, building on Ferrara et al. (2010). Despite their differences, there are important lessons from the current theory and practice of forecasting relevant to nowcasting, so we first reconsider forecasting, once the orphan of econometrics, but now a topic of major research, with extensive treatments in the four recent *Handbooks* by Clements and Hendry (2002), Elliott et al. (2006), Rapach and Wohar (2008), and Clements and Hendry (2011).

We now illustrate the inaccuracy of some previous flash estimates, and the frequency and magnitude of location shifts, using examples for UK GDP outturns. First, initial and later estimates often differ substantively, as seen in Figure 3.1.

Statistical agencies such as the UK Office for National Statistics (ONS) must base their flash estimates on incomplete and uncertain data. Hence, the need for nowcasts which combine both actual data for components that are known and forecasts of components that are unknown. Often surveys are used to 'infill' the unknown components to arrive at the flash estimate. This is a reasonable strategy when the surveys are good leading indicators, but such a methodology can prove disastrous in times of structural change. The *Financial Times* (25 July 2009) reported that 'the ONS said that its estimates were even less reliable than normal because the economy's unpredictability meant its models had "broken down".'[2] Any nowcasting strategy must be robust to structural breaks, requiring a methodology that can both detect structural breaks and rapidly adapt when such breaks occur. It is precisely at the point at which accurate nowcasts are most needed that current nowcasting methods are breaking down.

Secondly, unanticipated location shifts seem to occur all too often. As one of many possible examples, Barrell (2001) discusses six episodes of structural change in the 1990s, with other cases documented by Stock and Watson (1996) and Barrell et al. (1998). Also, Clements and Hendry (2001b) note the historical prevalence of failure in output forecasts for the UK, and the association of poor forecasts with major 'economic events'.

[2] http://www.ft.com/cms/s/0/c5529f00-7886-11de-bb06-00144feabdc0.html

Figure 3.1: UK GDP flash estimates and outturns.
Source: Bank of England Quarterly Bulletin, Autumn 2005, p.338.

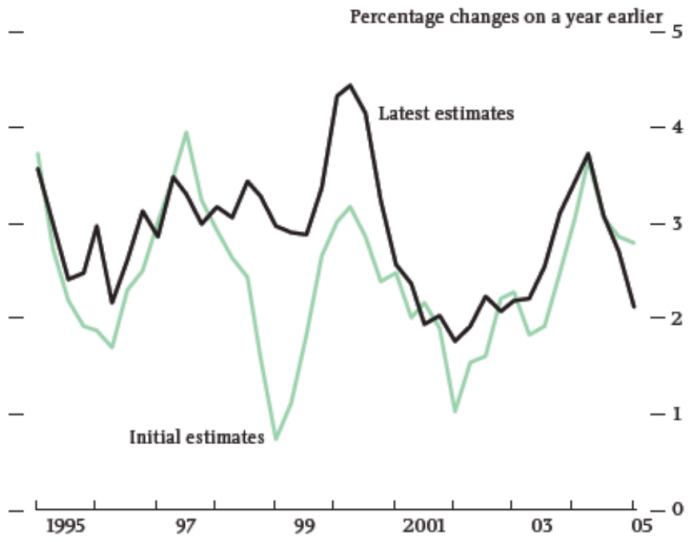
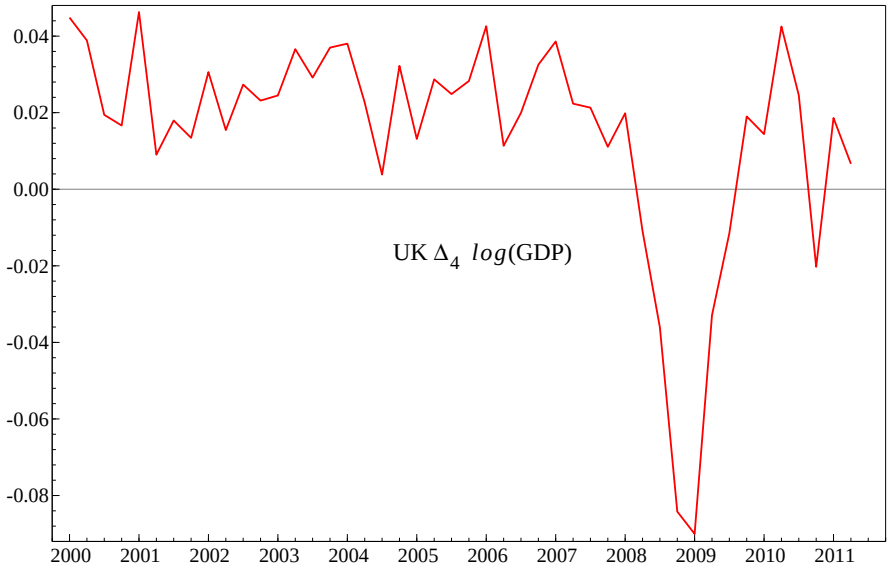


Figure 3.2 records the recently measured changes in the log of seasonally-adjusted UK GDP at annual rates over 2008(1)–2011(2) highlighting the sharp drop during 2008.

Figure 3.2: UK GDP year-on-year changes over 2000(1)–2011(2); data from June 2012.



The large empirical forecasting competitions, such as Makridakis et al. (1982) and Makridakis and Hibon (2000), reviewed respectively by Fildes and Makridakis (1995) and Clements and Hendry (2001a), produce results across many models on numerous time series for several horizons. Their general findings are consistent with the implications of unanticipated location shifts being the major source of forecast failures, with other structural breaks being much less pernicious: see Hendry (2000). Thus, we consider the reasons why forecast failure occurs and the implications for how to produce robust forecasts in order to apply those lessons to nowcasting.

3.1.1 Chapter structure

The structure of the chapter is as follows. Section 3.2 surveys a number of extant nowcasting approaches as background. Section 3.3 then reviews some of the fundamental issues that affect both forecasting and nowcasting.

Section 3.4 reviews the theory of forecasting in the face of breaks, and seeks to explain forecast failure, and its absence, as many of the lessons learned there are germane to nowcasting. Readers familiar with that literature can go directly to section 3.5. Otherwise, in section 3.4.1, we review the ten key difficulties confronting forecasting, using a scalar autoregressive-distributed lag (ADL) equation where the unmodeled variable is known into the future. Despite its simplicity, the illustration highlights all the main problems. section 3.4.2 seeks to understand the resulting rather strange set of forecast errors, and shows that the unique explanation is the presence or absence of location shifts, namely shifts in the long-run mean of the variable being forecast. When such shifts occur, forecast failure results irrespective of the goodness or otherwise of the in-sample model; when they do not occur, there is no obvious failure irrespective of how many parameters of the data-generating process (DGP) shift, and by how much. section 3.4.3 illustrates that robust forecasting devices may forecast better than any structural model in such shifting processes, as measured by root mean-square forecast errors (RMSFEs). section 3.4.4 extends robust forecasting devices to first-order integrated processes (denoted I(1)). Section 3.4.5 provides three illustrative empirical applications of *ex post* forecasting, to Japanese exports over 2008(7)–2011(6), when they fell by more that 70% year on year, and UK GDP over two recessions.

The huge number of variables potentially involved raises the need for automatic model selection when there are more variables than observations, so section 3.5 describes the basis of automatic model selection, section 3.5.1 outlines how *Autometrics* functions, section 3.5.2 discusses the role of mis-specification testing, with the special but important case of impulse-indicator saturation (IIS) described in section 3.5.3. Readers familiar with that literature can go directly to section 3.6, which considers using such findings to help guide the production of 'good' nowcasts.

All of these considerations apply to nowcasting an aggregate directly, as well as from disaggregates, which is the topic of section 3.6. Then section 3.6.1 considers aggregation and forecasting aggregates from disaggregates given the ever-changing database of observed disaggregates, with subsections 3.6.2 on common features and the role factor models might play, section 3.6.2 on the impacts of changing collinearity, section 3.6.3 on using all available information including section 3.6.3 on surveys, section 3.6.3 on covariate information, including Google query data in section 3.6.3, prediction market data in section 3.6.3 and seasonal adjustment in section 3.6.4. Next section 3.6.5 considers specifying, selecting and estimating forecasting models for all disaggregates. Section 3.6.6 formulates the nowcasting strategy, where comparisons with early observed series suggests using any major discrepancies to adjust the forecasts of missing related disaggregates before calculating the preliminary estimates of the aggregate. section 3.6.8 reviews the very practical problems of measurement errors and location shifts, and the role of judgment.

Section 3.7 applies the approach to *ex post* nowcasting UK GDP growth. section 3.7.1 discusses the data, with section 3.7.2 addressing the leading indicators considered. section 3.7.3 explains our nowcasting methodology, section 3.7.4 the actual specifications considered, with section 3.7.5–section 3.7.7 reporting the *ex post* 'real-time' nowcasting results. Section 3.7.8 investigates the impacts and handling of contemporaneous breaks. Section 3.8 draws some conclusions.

3.2 Nowcasting approaches

In stable environments, delayed releases of preliminary aggregates may nevertheless still provide sufficiently timely and accurate indicators of the present economic state. For instance, the Bureau of Economic Analysis (BEA) produced preliminary estimates of quarterly US GDP that showed the correct direction of growth 98% of the time before 2008: see Fixler and Grimm (2008). However, in rapidly changing environments, preliminary

releases may be both inaccurate and overdue, undergoing substantial revisions before the measurements are finalized, so can be poor approximations of the actual state of an economy.

The UK recession of 2008 provided an illustration of late and inaccurate reporting of dramatic macroeconomic developments. The first three preliminary estimates for the second quarter of 2008 produced by the ONS showed positive growth, when output had in fact dropped by 1.5%, a figure that was not reported until subsequent revisions. While national agencies may only get seriously wrong initial estimates infrequently, when they do so, such mistakes can have important policy consequences. This underlines the need to produce alternative accurate real-time estimates of GDP growth, based on statistical nowcasting models using extensive information sets.

The majority of quantitative methodologies of nowcasting are based on projecting GDP growth onto a set of explanatory variables linked to aggregate output as in Bánbura et al. (2011). Relevant information sets often contain monthly or higher frequency indicators, involving both mixed-data frequencies and asynchronous releases, producing unbalanced panels with missing observations at the end of the sample, a problem often referred to as the 'ragged edge': see Wallis (1986).

Nowcasting can also be based on combining forecasts of disaggregated components. Lütkepohl (2006) provides a detailed survey of aggregate versus disaggregate forecasts, which is also applicable to nowcasting. When the data-generating process (DGP) is known, disaggregated forecasts are at least as precise as a direct approach. When the DGP is unknown a-priori, Hubrich (2005) finds that direct forecasts of Eurozone HICP inflation have superior one-year-ahead predictive power compared to a combination of disaggregated forecasts, but this may not be the case when forecasting other variables or in different countries. Consequently, it is useful to understand the differences between direct nowcasts and a combination of disaggregated nowcasts as considered in Hendry and Hubrich (2011).

Notation can become overly complicated when every aspect of the changing environment of disaggregates and external variables and different frequencies are all denoted. In this section, we focus on frequency, so denote quarterly GDP by y_{t_q} where $t_q = 1, 2, \ldots, T_q$ represents quarters, with T_q representing the nowcast origin. GDP measures on a monthly frequency are denoted by $y_{t_m} = y_{t_q}$ and correspond to the months $t_m = 3t_q$, so that GDP is observed in month $t_m = 3, 6, 9, \ldots, T_m$, where $T_m = 3T_q$. Leading monthly indicators are denoted by z_{t_m} for $t_m = 1, 2, \ldots, T_m$, where the number of indicators is n. For each month t_m, the $t_m \times n$ matrix of exogenous explanatory variables is denoted by $\mathbf{Z}_{t_m}^1 = (\mathbf{z}_{t_m}, \ldots, \mathbf{z}_1)$, where $\mathbf{z}_{t_m} = (z_{1,t_m}, \ldots, z_{n,t_m})'$ is a vector of observations for a particular period. Given such information, the following nowcasting approaches are feasible.

Direct nowcasts

The least demanding approach is to nowcast the aggregate variable y_{t_q} using only its own past $\mathbf{Y}_{T_q-1}^1 = y_{T_q-1}, \ldots, y_1$:

$$\widehat{y}_{T_q|T_q-1} = f\left(\mathbf{Y}_{T_q-1}^1\right) \tag{3.1}$$

Specifying an autoregression for quarterly GDP estimates is a forecasting exercise using limited information of low frequency, so it is impossible to refine nowcasts between two instances of GDP publications. The latter becomes feasible if (3.1) is augmented by monthly indicators. Such a set of variables may include hard data like industrial production, financial indicators, interest rates measured at a higher frequency, or soft data like confidence or business surveys. Mitchell et al. (2005), Ferrara et al. (2010) and Bánbura et al. (2011) *inter alia* find that both types of indicators are relevant for explaining GDP growth.

In an augmented model, y_{t_q} is projected onto the set of monthly indicators $\mathbf{Z}_{t_m}$. A subset of $\mathbf{Z}_{t_m}$ is available for some vintage before T_q, say $T_q - v$, where $3T_m \leq T_q - v \leq 3\,(T_m + 1)$ and v is the maximum number of months for which the explanatory variable is available earlier than GDP. Since monthly indicators are released asynchronously with varying lags, a vector $(z_{1,T_m}, \ldots, z_{n,T_m})'$ will have zero entries for the unreleased variables. Consequently $\mathbf{Z}_{T_m}^1$ has ragged edges towards the end of the sample. Predictors are also released before the first GDP estimate is available. Therefore, the projection of y_{T_q} is conditioned on all available infor-

mation and not just that contemporaneous to T_q. The expanding flow of new monthly releases means that v converges to 0 as time goes by, and so at each instance $T_q - v$, a new augmented model can be formulated conditional upon all available information:

$$\widehat{y}_{T_q|T_q-v} = f\left(\mathbf{Y}^1_{T_q-1}, \mathbf{Z}^1_{T_q-v}\right) \tag{3.2}$$

Indirect nowcasts using disaggregates information

The aggregate y_{T_q} is defined as a weighted sum of N_T disaggregated components y_{i,T_q} as in:

$$y_{T_q} = \sum_{i=1}^{N_T} w_i y_{i,T_q},$$

where we ignore changes in weights. It is therefore possible to nowcast GDP as a weighted sum of individual component nowcasts. Assume that J_T out of N_T variables become available contemporaneously at time T_q, while $N_T - J_T$ disaggregates are only known up until lag $T - v^*$, where $v^* \geq v$, so not later than some of the indicators. The unknown disaggregates for $J_T + 1, \ldots, N_T$ are then nowcast using all available information at a disaggregated level:

$$\widetilde{y}_{i,T_q|T_q-v} = f\left(\mathbf{y}^1_{i,T_q-v^*}, \mathbf{Z}^1_{T_q-v}\right) \tag{3.3}$$

where $\widetilde{y}_{i,T_q|T_q-v}$ is the conditional nowcast and $\mathbf{y}^1_{i,T_q-v^*}$ are past observations for the component i. A combination of known and nowcast disaggregates yields the prediction of the aggregate:

$$\widetilde{y}_{T_q|T_q} = \sum_{i=1}^{J_T} w_i y_{i,T_q} + \sum_{i=J_T+1}^{N_T} w_i \widetilde{y}_{i,T|T_q-v} \tag{3.4}$$

Note that the aggregate nowcast $\widetilde{y}_{T_q|T_q}$ is contingent on a set of known components being released, which typically happens at the same time as the publication of the aggregate estimate. Consequently, nowcasting from (3.4) is possible at T_q, unless all components are unknown, needing nowcast at $T_q - v$.

Nowcasts using aggregated and disaggregated information

Two possibilities are available. To avoid the ragged-edge problem, nowcasts are produced from a balanced panel of both aggregated and disaggregated variables. A direct nowcast is formulated at $T - v^*$, when the last disaggregate is observed:

$$\bar{y}_{T_q|T_q-v^*} = f\left(\mathbf{Y}^1_{T_q-v^*}, \mathbf{y}^1_{i,T_q-v^*}, \mathbf{Z}^1_{T_q-v^*}\right) \tag{3.5}$$

This approach is a simplification since it disregards discrepancies in the publication times between aggregate and disaggregate information. This may lead to exclusion of important dynamics for the variables available at a later period, but not yet included in the information set. Alternatively, it is possible to perform a quasi-in-sample nowcast by utilizing all available data from an unbalanced panel:

$$\bar{\bar{y}}_{T_q|T_q} = f\left(\mathbf{Y}^1_{T_q-1}; \mathbf{y}^1_{1,T_q}, \ldots, \mathbf{y}^1_{J_T,T_q}, \mathbf{y}^1_{J_T+1,T_q-v^*}, \ldots, \mathbf{y}^1_{N_T,T_q-v^*}; \mathbf{Z}^1_{T_q-v}\right) \tag{3.6}$$

Specification (3.6) is a generalization of the representation with ragged edges where the aggregate, disaggregate and leading indicators information is included and uses the most complete data set. A test of the four approaches is conducted in Hendry and Hubrich (2011), who compare direct aggregate forecasting with a combination of disaggregate forecasts using a Monte Carlo simulation as well as an empirical study of the US consumer price index, while conditioning on both types of information. They conclude that direct forecasts are more accurate than combinations of disaggregate forecasts. It is also found that adding interdependent disaggregated variables with different stochastic structures improves nowcasts of the aggregate for both ap-

proaches when a parsimonious specification is estimated.

Most nowcasting methodologies utilize all types of available information when nowcasting, and try to find ways of dealing with the ragged-edge problem, rather than simplifying the analysis to balanced panels.

3.2.1 The 'in-filling' approach

Clements and Hendry (2003) reviewed the methodology used for nowcasting at the UK Office for National Statistics (ONS). The ONS then viewed forecasting any missing disaggregate data, or the aggregate directly, as a last resort for 'plugging gaps' when measured data were unavailable. 'In-filling' of missing disaggregates was based on either a version of the Holt–Winters exponentially-weighted moving average (EWMA) proposed by Holt (1957) and Winters (1960), or an autoregressive-integrated moving average (ARIMA) model following Box and Jenkins (1976). The ONS guidelines as to which to use emphasized whether the series was sufficiently important to merit going through an ARIMA model-fitting exercise, and whether time and expertise were available to do so, and the implementation of the chosen method was reviewed annually.

Both EWMA and ARIMA are well designed for handling first-order integrated (I(1)) data generated by a random walk subject to measurement errors. However, that does not seem to be an optimal assumption about the process generating the disaggregates in an economy subject to intermittent and often unanticipated location shifts, as neither method is robust after such breaks. Moreover, knowing that shifts have occurred in some of the measured variables suggests that a similar problem may affect related unobserved series. Multivariate approaches, 'outside' explanatory variables (other than as proxies), and mixed-frequency data suggested by e.g., Montgomery et al. (1998), were not used to obtain more accurate nowcasts. Thus, improvements to in-filling seem feasible.

3.2.2 Mixed-data sampling models

Introduced in Ghysels et al. (2004), the mixed-data sampling (MIDAS) approach links a low-frequency variable with selected predictors at higher frequency using a parsimonious restricted lag polynomial. Ghysels et al. (2007) and Andreou et al. (2011) provide comprehensive reviews of restrictions. The most common specification with coefficients of the lagged predictors being modeled as distributed lag functions is presented below. For simplicity of exposition, we consider a single indicator $z_{t_m} \in \mathbf{Z}_{t_m}$ explanatory variable. A MIDAS regression for a quarterly nowcasting horizon of $h_q = 3h_m$ quarters ahead, where h_m denotes the corresponding monthly horizon, is then given by:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 B\left(L_m;\theta\right) z^{(3)}_{t_m+v} + \epsilon_{t_m+h_m} \tag{3.7}$$

where $B\left(L^m;\theta\right)$ is the exponential Almon lag function with:

$$B\left(L^m;\theta\right) = \sum_{k=0}^{K} B\left(k;\theta\right) L_m^k, \qquad B(k;\theta) = \frac{exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=0}^{K} exp(\theta_1 k + \theta_2 k^2)} \tag{3.8}$$

where $L_m^k z_{t_m} = z_{t_m-k}$ is the monthly lag operator, and $\theta = \{\theta_1, \theta_2\}$ is the parameter vector. The superscript (3) in $z^{(3)}_{t_m+v}$ denotes that the predictors are observed with frequency three times as high as the dependent variable in the same time frame. Finally, K is the longest lag of the explanatory variable allowed to enter the regression. Since β_1 captures the overall effect of $z^{(3)}_{t_m+v}$ and its lags via $y_{t_m+h_m}$, identification of the coefficient is ensured by normalization: $\sum_{k=0}^{K} B\left(k;\theta\right) = 1$. Given the estimated parameters $\widehat{\beta}_0, \widehat{\beta}_1$ and $\widehat{\theta}$, the nowcast at $T_m + v$ is computed as:

$$y_{T_q+h_q|T_m+v} = \widehat{\beta}_0 + \widehat{\beta}_1 B\left(L_m;\widehat{\theta}\right) z^{(3)}_{T_m+v} \tag{3.9}$$

An extension of this basic model, proposed in Ghysels et al. (2004), considers the autoregressive representation:

$$y_{t_q+h_q} = \beta_0 + \lambda y_{t_m} + \beta_1 B\left(L_m;\theta\right)\left(1-\lambda L_m^3\right) z_{t_m+v}^{(3)} + \epsilon_{t_m+h_m} \tag{3.10}$$

where the autoregressive coefficient λ is estimated jointly with the lag polynomial parameters and the regression coefficients using non-linear least squares. An empirical implementation of an autoregressive MIDAS model is presented in Clements and Galvão (2008), who use a small subset of indicators that may be relevant in explaining GDP growth. They find that industrial production and capacity utilization significantly reduce the root mean-square error (RMSE) for US GDP nowcasts as compared to an autoregressive benchmark, similar to that in (3.1). In a subsequent application of the MIDAS approach, Clements and Galvão (2009) find that single indicator models are not robust predictors of recessions, but multiple-indicator MIDAS models improve nowcasting accuracy relative to a forecast combination, especially when predicting the direction of growth. This suggests using extensive information to improve nowcasting performance, but estimation of multiple-variable MIDAS models is susceptible to mis-specification. The next model attempts to resolve that issue by summarizing all variation in the explanatory variables using a smaller number of combinations.

3.2.3 Factor models

The so-called 'curse of dimensionality' seems relevant in nowcasting since the information set may contain a large number of explanatory variables. Estimating all parameters may be impossible (an issue addressed in section 3.5.3 below), whereas omitting relevant variables will result in mis-specification and reduced explanatory power. One solution to both these problems is to summarize the dynamics of the monthly indicators using a small set of common factors and apply those as explanatory variables. Boivin and Ng (2005) provide a summary of factor models for single frequency data. A general approach includes two steps. Initially the factors are computed, followed by an estimation of a regression for the dependent variable augmented with these factors.

A dynamic factor representation of a vector $\mathbf{x}_t$ has the following form:

$$x_{i,t} = \lambda_i(L)'\mathbf{f}_t + \xi_{i,t} \tag{3.11}$$

where $\mathbf{f}_t$ is an $\bar{r}\times 1$ vector of dynamic common factors, $\lambda_i(L)$ is a lag polynomial of degree s, L is the lag operator, and ξ_t is a $n\times 1$ vector of idiosyncratic components. Given this representation, Stock and Watson (2002b) propose to stack the factors in an $r\times 1$ vector $\mathbf{F}_t = \left(\mathbf{f}_t', \ldots, \mathbf{f}_{t-s}'\right)'$ where $r \leq \bar{r}(s+1)$ to obtain a static representation of a factor model:

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \xi_t \tag{3.12}$$

where $\mathbf{\Lambda}$ is a matrix of loadings, with the ith row of $(\lambda_{i,0},\ldots,\lambda_{i,s})$, of the common static factors $\mathbf{F}_t$. In both representations the idiosyncratic component is orthogonal to the factors. The common components are computed as:

$$\chi_t = \mathbf{\Lambda}\mathbf{F}_t \tag{3.13}$$

In the classical case with no ragged edges, a balanced panel $\mathbf{Z}_{T_m}^1$ does not have have any missing observations at the end of the sample. In this case, there are two main estimation procedures considered in the literature. A classical approach proposed by Stock and Watson (2002a) and Stock and Watson (2002b) uses static principal components (PCs) of the sample covariance matrix. Principal components $\widehat{\mathbf{F}}$ of a $T\times n$ matrix of the explanatory variables are calculated as $\widehat{\mathbf{F}} = \mathbf{X}\widehat{\mathbf{H}}$ where $\widehat{\mathbf{H}} = (\widehat{\mathbf{h}}_1,\ldots,\widehat{\mathbf{h}}_n)$ is the matrix of eigenvectors corresponding to the largest eigenvalues $(\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n)$, of the covariance matrix $\widehat{\mathbf{C}}$ of $\mathbf{X}$, given by the eigenvalue decomposition:

$$\widehat{\mathbf{C}} = T^{-1}\mathbf{X}'\mathbf{X} = \widehat{\mathbf{H}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{H}}'.$$

Dimensionality reduction is obtained by selecting a few factors corresponding to the largest eigenvalues that explain most of the variation in the data on $\mathbf{X}$. In the second stage, an OLS projection of the dependent variable onto a set spanned by the factors $\widehat{\mathbf{F}}$ is estimated.

As an alternative to static PCs, Forni et al. (2001) consider an approach of estimating dynamic factors using generalized PCs where the weight of each observation is proportional to its signal-to-noise ratio. They use non-parametric techniques to forecast the factors taking into consideration restrictions on the dynamic factor structure. Forni et al. (2005) argue that this method provides theoretical efficiency improvements as compared to a static factors representation.

Since the introduction of the static factor model, there have been various empirical studies confirming its superiority to models comprising a small number of variables: see Stock and Watson (2002b), Camacho and Sancho (2003), Brisson et al. (2003) and Artis et al. (2005) inter alia. However, its performance has also been questioned in Giacomini and White (2006), who show that in a moving-window framework, static factors do not necessarily provide a forecasting accuracy improvement. Moreover, Banerjee et al. (2005) for Euro-area inflation and GDP, and Schumacher and Breitung (2008) for German data, show that static factors need not perform better than single indicator models. Castle, Clements, and Hendry (2013) apply the approach in Castle, Doornik, and Hendry (2013) (who model using all the variables jointly with their PCs) to forecasting US GNP, but find that when the sample is extended over the recent recession, neither variables nor PCs (nor both) perform well.

Thus, despite various empirical and simulation based comparisons, there is no consensus on whether static or dynamic factor models forecast better. Forni et al. (2003), who analyze financial time series linked to Euro-area inflation and GDP, D'Agostino and Giannone (2012) using a large cross-sectional data study for the US, and a survey in Stock and Watson (2006), all conclude that forecasting accuracy of the two specifications is similar so that it is difficult to distinguish a clear-cut winner. In particular, for the original US data set used in Stock and Watson (2002a), only a few factors provide substantial additional predictive power as compared to a simple autoregressive specification for forecasting of inflation and industrial production, and dynamic factors do not yield any improvement in predictive accuracy.

Conversely, supporting the empirical superiority of dynamic factors are the studies by den Reijer (2005) for the Dutch economy, Schumacher and Breitung (2007) for German data, and D'Agostino et al. (2013) for Irish GDP, who find substantial accuracy improvement when using the approach in Forni et al. (2001). In the opposite camp is the evidence provided by Boivin and Ng (2005), who find that the static approach is robust to model mis-specification and performs better when forecasting variables with unknown dynamic structure, since it imposes no restrictions on the factors and forecasts the series directly. Overall, both approaches tend to achieve accuracy improvement as compared with benchmark models. Since there is no clear leader in this race, and much room for further investigation as data properties change over time, the static factor approach is slightly preferred for its practical simplicity.

3.2.4 Factors with ragged edges

In the context of nowcasting, factor estimation is more problematic in view of the missing observations at the end of the sample. The ragged-edge problem makes the panel unbalanced, so prevents direct estimation of static or dynamic factors using the methods just discussed. Therefore a procedure that can handle missing data is required.

A simple approach is discussed in Altissimo et al. (2010), who propose to use vertical re-alignment of the data. In practice, this implies that the most recent available data point is assigned to be the current value of the series. For instance, suppose that at time T_q some indicator z_i is available with a lag of l_i, so the most recently released value is z_{i,T_q-l_i}. Vertical re-alignment assigns this value to z_i at period T_q:

$$\widetilde{z}_{i,T_q} = z_{i,T_q-l_i} \tag{3.14}$$

When applied to all series for $t_l = l_i + 1, \ldots, T_q$, given that l_i is the maximum release lag, this procedure results in a balanced panel $\widetilde{\mathbf{Z}}^1_{T_q}$. Despite being undemanding computationally, and for this reason convenient practically, this method distorts the dynamic correlation structure each time realignment is applied, since the data are released at different times and various revisions are applied: see Marcellino and Schumacher (2010).

An alternative way of constructing a balanced panel from data with missing observations is, as suggested by Stock and Watson (2002b), to use an Expectation-Maximization (EM) algorithm for the static factors in (3.12). The algorithm constructs a balanced panel through a recursive procedure that updates the missing observations and re-estimates the factors at the last column of the data matrix that contains zeros for the unobserved values. Angelini et al. (2006) point out that although there are more sophisticated methods of data interpolation and backdating, such as large factor state-space representations, as in Kapetanios and Marcellino (2010), or utilizing the Kalman Filter as in Doz et al. (2011), the EM algorithm performs interpolation and estimation of the factors consistently and is computationally convenient. Given that many of the most recent studies, including D'Agostino and Giannone (2012), Schumacher and Breitung (2007), Schumacher and Breitung (2008) and Marcellino and Schumacher (2010) find that when correcting for the ragged-edges in nowcasing, the EM algorithm does not perform worse than the more computationally demanding methods, it may be preferred for its practical convenience.

In an empirical study of factor models, Bánbura et al. (2011) apply dynamic factors to nowcasting Euro-area GDP, and find that as new high-frequency information arrives throughout the reference quarter, nowcasting accuracy improves monotonically. Less timely hard data variables become important closer to the end of the quarter, while survey data provides an early indication for GDP growth. This is confirmed in Giannone et al. (2009) and Bańbura and Rünstler (2011) who find that survey data provide a good early estimate of GDP, with real activity and financial data variables playing a relatively less important role. The same conclusion is reached by Giannone et al. (2008), who use dynamic factor models with the Kalman smoother updates for missing observations in a pseudo real-time US GDP nowcasting exercise. These findings provide additional justification for utilizing mixed frequency ragged-edged data of varying types in nowcasting as opposed to formulating balanced quarterly forecasting models.

The main difficulties for factor-based nowcasts are the ragged-edge problem and mixed data frequencies. As an alternative to using factors directly as explanatory variables, a hybrid MIDAS-factor model could be formulated. This approach combines all available information in a parsimonious way, thereby resolving the problems of mixed frequencies, ragged edges and dimensionality. In an empirical study, Marcellino and Schumacher (2010) and Kuzin et al. (2009) use a MIDAS-factor approach to nowcast GDP growth in six countries. For $r > 1$, they estimate common factors $\widehat{\mathbf{F}}_{t_m} = \left(\mathbf{f}'_{1,t_m}, \ldots, \mathbf{f}'_{r,t_m}\right)$ and use them in the following MIDAS regression:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \sum_{i=1}^{r} \beta_{1,i} B_i\left(L_m; \theta_i\right) \widehat{\mathbf{F}}_{t_m}^{(3)} + \epsilon_{t_m+h_m} \tag{3.15}$$

where $B_i\left(L_m; \theta_i\right)$ are the Almon lag polynomials as in (3.8). Kuzin et al. (2009) provide a comparison of the ragged-edge correcting technique within the MIDAS and MIDAS-factor frameworks. They find that all specifications have similar nowcasting performance, and that using real-time ragged-edge information for nowcasting outperforms balanced panels. However, their results suggest that pooling single-indicator models tends to outperform factor models. Eickmeier and Ziegler (2008) reach the same conclusion based on US and UK data. This effect is robust to the selection of the pooling weighting schemes. Kuzin et al. (2009) also conduct sequential model selection, and conclude that the results do not match the accuracy achieved by pooling. Their selection procedure is, however, based on a single simple criterion, and more sophisticated model selection techniques could well improve on this benchmark case. Another alternative to factor models and MIDAS regressions that assures dimension reduction and parsimonious model representation is provided by Bayesian methods. However, following De Mol et al. (2008) who compare Bayesian shrinkage techniques with factor models and find that the two are on par, discussion of the former is omitted here: a simpler and intuitively convenient nowcasting approach is described in the next sub-section.

3.2.5 Bridge equations

MIDAS regressions and factor models are parsimonious but restrictive ways of combining extensive informations set with mixed frequency data and ragged edges. Moreover, their results are hard to interpret: the lag distribution of MIDAS coefficients is ad-hoc, and common factors may not be uniquely identifiable. PCs are a

transformation of the original data, which provides no additional information about the underlying link between the target series and exogenous variables. For these reasons, a slightly less restrictive approach is based on equations that construct a direct 'bridge' between aggregate measures and a set of explanatory variables. Bridge equations are popular in central banks, mainly due to their interpretability. The approach specifies an equation for GDP growth, Δy_{t_q}, at a quarterly frequency, as in Mitchell (2009):

$$\Delta y_{t_q} = c + \sum_{i=1}^{p} \alpha_i \Delta y_{t_q - i} + \sum_{j=0}^{p} \sum_{i=1}^{k} \beta_{ij} z_{i,t_q - j} + u_{t_q} \text{ for } t = 1, \ldots, T \tag{3.16}$$

where k quarterly indicators z_{i,t_q} are transformed to remove unit roots, p is the lag length and u_t are assumed to be IID residuals. The regression parameters describe the link between the leading indicators and GDP growth, a feature missing in both the MIDAS and factor models.

The quarterly indicators of z_{i,t_q} in (3.16) are transformed versions of the observed monthly indicators, a subset of which is unobserved in-sample due to the ragged-edge problem. The individual equations that forecast the missing values of the monthly indicators are then specified to correct for the unbalanced panels. In the simplest setting, a univariate AR(p) process could be used:

$$x_{i,t_m} = \sum_{i=1}^{p} \gamma_i x_{i,t_m - i} + e_{i,t_m} \tag{3.17}$$

where e_{i,t_m} is assumed to be white noise. The quarterly indicators are then formed using a combination of the observed monthly data, where available, and their forecasted values. Essentially, bridge equations provide an alternative tool for correcting for the ragged-edge problem by directly predicting the missing monthly observations required for nowcasting as in Mariano and Murasawa (2003). Also, Barhoumi et al. (2008) find that utilizing monthly information to forecast the missing data tends to improve nowcasts compared with balanced quarterly regressions. Bridge equations can also take the form of a factor model, as described in Gertler and Rogoff (2005) and Giannone et al. (2008), such that:

$$\Delta y_{t_q} = c + \delta' \mathbf{f}_{t_q} + u_{t_q} \tag{3.18}$$

where $\mathbf{f}_{t_q}$ are quarterly factors aggregated over their monthly representations $\mathbf{f}_{t_m}$ that are estimated from monthly data on the leading indicators using, for example, a dynamic representation:

$$x_{t_m} = \lambda' \mathbf{f}_{t_m} + \xi_{t_m}$$

as in (3.12), and the missing factors are forecast using the Kalman Filter. In an empirical application of bridge equations, Mitchell (2009) considers regression-based nowcasts for UK GDP with the information set containing industrial production indicators and qualitative survey data from the Confederation of British Industry (CBI). A simple factor model is estimated, linking the factors obtained using bridge equations, that can successfully nowcast the start of the recession in 2008 two months before the publication of the official data by the ONS. In a similar fashion, Kitchen and Monaco (2003) and Rünstler and Sédillot (2003) apply bridge equations to US and Euro-area data respectively, and show that nowcasting improves on naive autoregressive and random-walk models.

Diron (2008) uses hard data, surveys and sentiment indicators to nowcast Euro-area GDP growth, and shows that nowcasts based on pseudo real-time forecasts of monthly indicators using bridge equations differ to nowcasts obtained using revised data. However, revisions to GDP and monthly indicators constitute only a small fraction of the nowcasting error. This may not be the case for the UK, since the ONS failed to measure a decline in output in the second quarter of 2008 at the time it actually happened.

Given that most studies concentrate on relatively stable periods with no major macroeconomic perturbations, the finding in Diron (2008) provides reassurance that nowcasting methodologies can improve not only on simple benchmark models, but also perform better than the measurements by the statistical agencies to yield

more timely and accurate estimates. This is confirmed by the ex-post nowcasting exercise in Mitchell (2009) who essentially predicted the recession in the UK half a year before the ONS reported a decline in output. Moreover, the ONS estimate was published only after an additional quarter had elapsed. These findings suggest that bridge equations provide an easy, interpretable framework for nowcasting that performs as well as more sophisticated and computationally demanding models. Raw implementation of bridge equations in large information sets is, however, subject to a dimensionality curse, since the number of parameters to be estimated grows linearly with the number of monthly indicators n. A feasible solution to that problem is to use model selection, as discussed in section 3.5.

We now consider the difficult issue of how location shifts affect general model formulation.

3.3 Some fundamentals

In this section, we show the key result that when unanticipated location shifts occur, conditional expectations of a future variable need not be unbiased even based on the in-sample DGP and given all available information. Thus, other devices for nowcasting and forecasting may outperform.

The theory of economic forecasting when an econometric model coincides with a stationary DGP has been well developed along lines first proposed in Haavelmo (1944).

Consider an $n \times 1$ vector $\mathbf{x}_t \sim \mathsf{D}_{\mathbf{x}_t}(\mathbf{x}_t|\mathbf{X}_{t-1}, \theta)$ for parameters $\theta \in \mathbf{\Theta} \subseteq \mathcal{R}^k$, where $\mathsf{D}_{\mathbf{x}_t}(\cdot)$ denotes the data density of $\mathbf{x}_t$, conditional on $\mathbf{X}_{t-1} = (\dots \mathbf{x}_1 \dots \mathbf{x}_{t-1})$. A forecast $\widetilde{\mathbf{x}}_{T+h|T} = \mathbf{f}_h(\mathbf{X}_T)$ is desired at T for $\mathbf{x}_{T+h}$ in time $T+h$ using all the available information. The main issue is how to select $\mathbf{f}_h(\cdot)$, and the well-known answer is the conditional expectation, $\widehat{\mathbf{x}}_{T+h|T} = \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T]$ which is unbiased:

$$\mathsf{E}\left[\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right) \mid \mathbf{X}_T\right] = \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T] - \mathsf{E}[\mathbf{x}_{T+h}|\mathbf{X}_T] = \mathbf{0} \tag{3.19}$$

and $\widehat{\mathbf{x}}_{T+h|T}$ has the smallest mean-square forecast-error matrix of unbiased predictors:

$$\mathsf{M}\left[\widehat{\mathbf{x}}_{T+h|T} \mid \mathbf{X}_T\right] = \mathsf{E}\left[\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)' \mid \mathbf{X}_T\right] \tag{3.20}$$

Such results sit uneasily with the intermittent occurrence of forecast failure based on models that seek to capture conditional expectations. Either econometric models fail badly in their aim, or the results in (3.19) and (3.20) are not empirically relevant. Since the DGP is not known, the former is quite possible: there are many problems learning about $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$ and θ, involving the specification of the set of relevant variables $\{\mathbf{x}_t\}$, measurement of those $\mathbf{x}$s, the formulation of $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$, modeling the relationships between the variables, and estimating θ, all of which introduce in-sample uncertainties. The unknown properties of $\mathsf{D}_{\mathbf{X}_{T+H}^{T+1}}(\cdot)$ determine the forecast uncertainty, which grows as H increases, especially for integrated data. Thus, the conventional answer for the failure of (3.19) and (3.20) to hold in practice is that we fail to model conditional expectations well enough.

However, results from our own research provide a framework for explaining why forecast failure can occur even when the best structural model is used, which coincides with the DGP in-sample, and indeed even when the future values of all unmodeled variables are known before forecasting. Unanticipated location shifts have a profound implication: conditional expectations formed at time T from information available at that date for an outcome at $T+1$ are neither unbiased nor minimum variance when unanticipated location shifts occur: see Hendry and Mizon (2011b) and Hendry and Mizon (2014).[3] When the mean of the data distribution has shifted from T to $T+1$, the previous conditional expectation is not centered on the relevant new mean. Since many theorems about economic forecasting begin with the presumption that conditional expectations provide unbiased and minimum variance predictors, and hence deduce that econometricians should use the

[3] Also see http://www.voxeu.org/article/why-standard-macro-models-fail-crises for a recent explanation.

conditional expectations of models, a major rethink is required. Thus, unanticipated changes in $\mathsf{D}_{\mathbf{x}_{T+H}^{T+1}}(\cdot)$ and θ are the culprit. It is manifest that similar implications hold for nowcasts.

Next, rather than being a pernicious necessity to be avoided at all possible costs, model selection by reduction in a general-to-specific (*Gets*) approach is usually beneficial. A number of studies have demonstrated that unless the prior theory is precisely correct, selection delivers improved models: see Doornik (2009a), Doornik and Hendry (2013), Castle, Doornik, and Hendry (2011), and Castle and Hendry (2014a). If the theory model is correct, then even after considerable search, selection can still deliver identical estimates for its parameters to those obtained by direct fitting: see Hendry and Johansen (2015). By using an extended *Gets* approach, there can even be more candidate variables to be selected from than the number of observations: see Hendry and Krolzig (2005), Doornik (2007) and the extensive overview in Hendry and Doornik (2014). In turn, such a capability makes it possible to check for outliers and location shifts at every observation during selection using impulse-indicator saturation (IIS) discussed by Hendry et al. (2008), and Johansen and Nielsen (2009), and most recently step-indicator saturation (SIS): see Doornik and Hendry (2013), Castle, Doornik, Hendry, and Pretis (2015), and Ericsson (2012). Castle, Doornik, and Hendry (2012) investigate the performance of IIS in detecting multiple location shifts during selection, including breaks close to the start and end of the sample (as well as correcting for non-normality), and Hendry and Mizon (2011a) provide an empirical application to demand for food in the US.

Finally, and a consequence of the previous findings, the quality of forecasts from a model depends both on how it is used and what the properties of the forecast period are, as much as on the specification of the model itself and how it was estimated: see e.g., Clements and Hendry (1998), Clements and Hendry (1999) and Hendry (2006). We structure our analysis of nowcasting around these forecasting issues.

In part, the increasing realization of unanticipated location shifts as a key problem when modeling and forecasting has necessitated an emphasis on the invariants of the economic process, first discussed by Frisch (1938) and later highlighted by Lucas (1976), inducing the reformulation of the notion of exogeneity in Engle et al. (1983) and Engle and Hendry (1993). However, the problem of unanticipated shifts was central to the prescient work of Smith (1927) and Smith (1929), which somehow was forgotten despite being published just before the Great Depression: see Mills (2010) who focuses on Bradford Smith's research on the 'nonsense regressions' problem. Moreover, new tests of the key concepts of shifts and invariance have appeared, including several based on impulse-indicator saturation: see e.g., Hendry and Santos (2010). Thus, computational power and hugely improved algorithms have built on theoretical understanding to radically alter how econometric models can be constructed, and should be used.

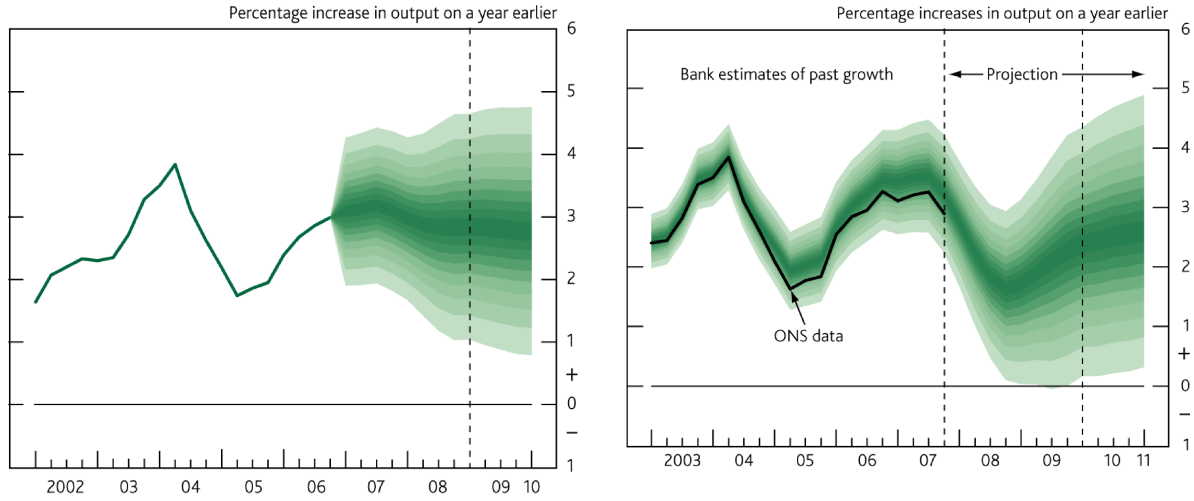
3.4 Explaining forecast failure–and its absence

In this section, we illustrate the key result that the presence or absence of location shifts is the main determinant of forecast failure, such that in open models with zero intercepts, even knowing the future values of unmodeled variables and using the in-sample DGP need not avoid forecast failure.[4] Model mis-specification and parameter estimation uncertainty play secondary roles, illustrated by robust forecasting devices performing better than a structural model.

Forecasting is difficult. The left hand panel of figure 3.3 from the *Bank of England Inflation Report* for February 2007 shows its forecasts at that time for annual changes in UK GDP: 'steady as she goes' through the end of 2009. The right hand panel of figure 3.3 updates their forecasts to February 2008 through the end of 2010: a distinct slowdown is now envisaged. However, it is nothing like the 8% fall at annual rates that materialized, which was unanticipated at that stage: Figure 3.2 recorded the actual changes in UK GDP at annual rates over 2008(1)–2011(2). As seen, there was a sharp drop at the end of 2008. In late 2006 and very early 2007, there was little to suggest the UK's 16 years of uninterrupted growth was about to end: even if the US subprime mortgage problem had been high on the Bank's radar, it is unlikely anyone foresaw it nearly bringing

[4] This section draws heavily on Hendry and Mizon (2012a) and is included with kind permission from Grayham E. Mizon.

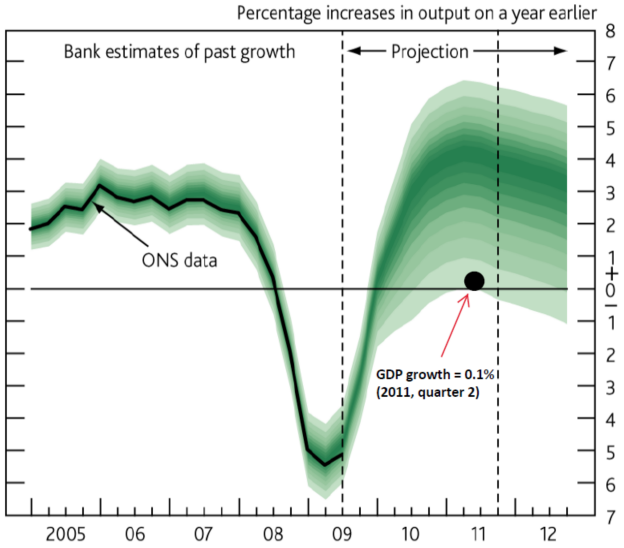
Figure 3.3: UK GDP projections based on market interest rate expectations. Left hand panel - February 2007; right hand panel - February 2008.
Source: Bank of England Inflation Report.



the whole world economy to its knees.

Even a more recent set of Bank forecasts reveal the continuing difficulty of accurate forecasts as figure 3.4 shows: the outcome in 2011(2) is at the lowest end of the range. Moreover, their uncertainty in the past outcomes is non-negligible.

Figure 3.4: GDP projections based on market interest rate expectations and £200 billion asset purchases.
Source: Bank of England Inflation Report, November 2009.



Such graphs emphasize that economies are non-stationary and evolving, where even the best economic model differs from the DGP, and forecast failures occur. Such failures are also associated with econometric model forecasts being out-performed by so-called 'naive devices', and that dates from the early history of econometrics, partly because almost no forecasting models allow for location shifts even after they occur, although such shifts clearly occur empirically. Note that adding variables which 'explain' shifts in-sample will improve forecasts only if their shifts can be forecast in turn, but again that need not entail that nowcasts cannot be improved. We now focus on 1-step ahead forecasts as the aim is to elucidate corresponding issues for nowcasting.

3.4.1 Scalar autoregressive-distributed lag example

Consider an autoregressive-distributed lag DGP with a known exogenous $\{z_t\}$:[5]

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right] \tag{3.21}$$

which is stationary in sample with all parameters (ρ, γ, σ_ϵ^2) constant and $|\rho| < 1$. When ρ, γ are known and x_T and z_{T+1} are observed without error, the optimal forecast for x_{T+1} is:

$$\widehat{x}_{T+1|T} = \mu + \rho x_T + \gamma z_{T+1} \tag{3.22}$$

producing an unbiased forecast:

$$\mathsf{E}\left[\left(x_{T+1} - \widehat{x}_{T+1|T}\right) \mid x_T, z_{T+1}\right] = \mathsf{E}\left[\mu - \mu + (\rho - \rho)\, x_T + (\gamma - \gamma)\, z_{T+1} + \epsilon_{T+1}\right] = 0$$

which is zero with the smallest possible conditional and unconditional variance determined by $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$:

$$\mathsf{V}\left[\left(x_{T+1} - \widehat{x}_{T+1|T}\right) \mid x_T, z_{T+1}\right] = \mathsf{V}\left[\epsilon_{T+1}\right] = \sigma_\epsilon^2.$$

Here, $\mathsf{D}_{\mathbf{X}_T^1}(\cdot)$ implies $\mathsf{D}_{\mathbf{X}_{T+1}^{T+1}}(\cdot)$ and $\mathsf{D}_{\mathbf{X}_{T+1}^{T+1}}(\cdot) = \mathsf{IN}\left[\mu + \rho x_T + \gamma z_{T+1}, \sigma_\epsilon^2\right]$. The focus of much of the literature on forecasting has been on the additional consequences of estimating the parameters of structural models like (3.21), often assumed to coincide with the DGP.

However, there is a long list of potential problems. These include that the specification may be incomplete if (e.g.) $\mathbf{x}_t$ is not a scalar; measurements might be incorrect if (e.g.) $\widetilde{\mathbf{x}}_t$ was observed, not $\mathbf{x}_t$; the formulation might be inadequate if (e.g.) an intercept was needed or z_t was omitted; the modeling might have gone wrong if (e.g.) x_{t-2} was incorrectly selected; estimating ρ adds a bias, $(\rho - \mathsf{E}[\widehat{\rho}])x_T$, and a variance $\mathsf{V}[\widehat{\rho}]x_T^2$ (as does estimating γ if z_t is included); the forecast analysis earlier in this section assumed $\epsilon_{T+1} \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$ but $\mathsf{V}[\epsilon_{T+1}]$ could differ; a multi-step forecast error $\sum_{h=1}^{H} \rho^{h-1}\epsilon_{T+h}$ would arise when $H > 1$, like that in the first Chart above, leading to $\mathsf{V}[\sum_{h=1}^{H} \rho^{h-1}\epsilon_{T+h}] = \frac{1-\rho^{2H}}{1-\rho^2}\sigma_\epsilon^2$ so if $\rho = 1$, the forecast variance would be trending as $H\sigma_\epsilon^2$; and if μ, ρ or γ changed, forecast failure could occur. Interactions between all of these difficulties could compound the problem for forecasters. Fortunately, as shown in Clements and Hendry (1999), most problems do not lead to the kind of mis-forecasting seen in the charts above, as we now illustrate. Unfortunately, as we now show, even after determining the source of such forecast failures, many serious difficulties remain for the structural modeler.

To simplify the first round of analysis, we set $\mu = 0$ in the DGP (3.21) where this is known. That enables us to contrast forecasts like (3.22) knowing (ρ, γ) with estimating those parameters from a sample of accurate data over $t = 1, \ldots, T$:

$$\widetilde{x}_{T+1|T} = \widehat{\rho} x_T + \widehat{\gamma} z_{T+1} \tag{3.23}$$

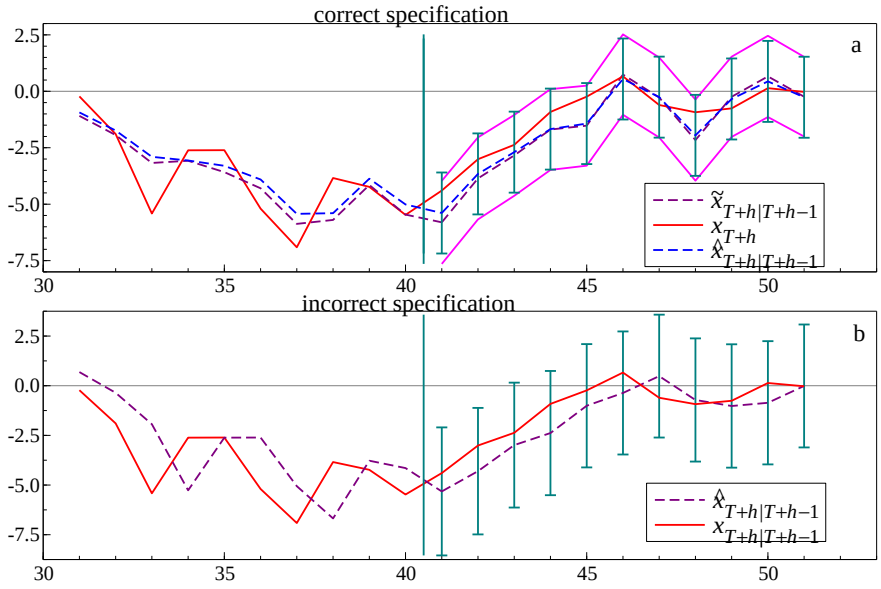
Next, we mis-specify the model by omitting z_t, then also change ρ. The graphs that follow are based on a single random draw of $T = 50$ observations from (3.21) with $z_t \sim \mathsf{IN}[0, 1]$ when $\rho = 0.8$, $\gamma = 1$, and $\sigma_\epsilon^2 = 1$. A later theoretical analysis will conform the results are general despite the specificity of the illustration, but the graphs highlight the key issues.

Figure 3.5 panel a, records the forecasts both when ρ and γ are known and constant, and when they are estimated. $\widehat{x}_{T+h|T+h-1}$ from (3.22) is shown with error bars of $\pm 2\widehat{\sigma}$, whereas forecasts when estimating (ρ, γ) are shown as $\widetilde{x}_{T+h|T+h-1}$ with forecast interval bands. As can be seen, the two sets of forecasts are almost identical, with only a small increase in uncertainty from estimation. Thus, estimation *per se* does not seem to be a major problem, and is well known to be of probabilistic order $\mathsf{O}_p(T^{-1})$.

Next, we consider the impact on forecast accuracy and precision of incorrect specification, here corresponding

[5] This example is taken from Hendry and Mizon (2012a), which should be cited as the original source, as its use here is to provide a complete analysis. We include a contemporaneous 'exogenous' variable to link more closely with nowcasting.

Figure 3.5: Artificial data forecasts under correct and incorrect specification.

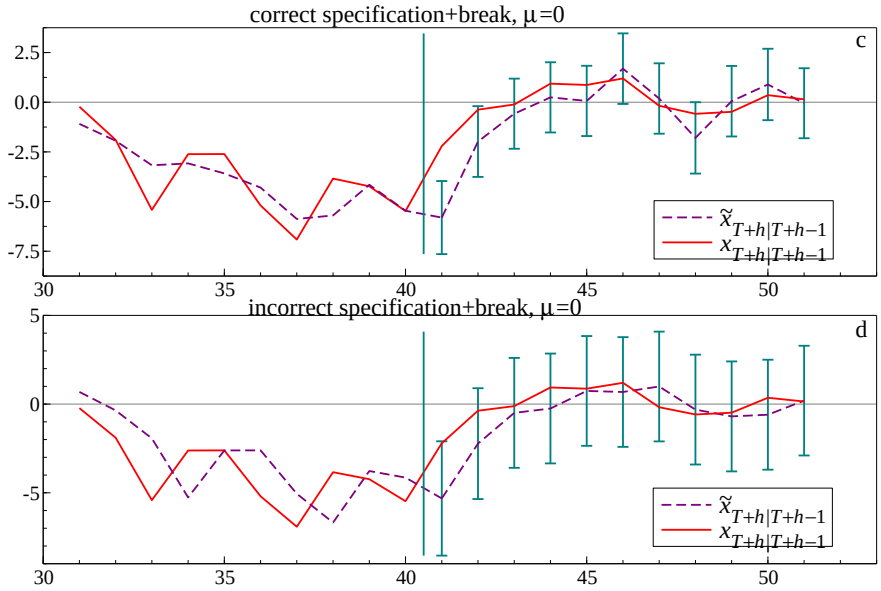


to inadvertently omitting z_t, shown in Figure 3.5 panel b. When z_t is omitted both in estimation and forecasting the resulting forecasts are:

$$\widetilde{x}_{T+1|T} = \widetilde{\rho} x_T \tag{3.24}$$

These are clearly poorer than under correct specification, but well within the *ex ante* forecast intervals shown. Thus, forecast failure is not just a mis-specification problem either.

Figure 3.6: Correct and incorrect specifications with changed ρ.



Finally in this group, at $T = 40$ we shift ρ to $\rho^* = 0.4$, then back to $\rho = 0.8$ at $T = 46$ till $T = 50$, so the DGP reverts to its previous state, mimicking a short, sharp regime switch in the dynamics, denoted:

$$x_{T+h} = \rho^* x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{3.25}$$

over $h = 1, \ldots, 5$, before again becoming:

$$x_{T+h} = \rho x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{3.26}$$

over $h = 6, \ldots, 10$. To allow for possible interactions with mis-specification, we first return to forecasting from (3.23), then also use (3.24). Figure 3.6 panel c reports the former: there is only a slight impact from halving ρ, then almost none from doubling it again, so forecast failure is not just a problem of changing parameters.

Next, we use (3.24), so all of estimation uncertainty, mis-specification, and non-constancy occur, yet again there is little noticeable impact from halving then doubling ρ as seen in figure 3.6 panel d. Whatever underlies outcomes as in the GDP Figure 3.2, those three mistakes do not explain it. Nor could reasonable measurement errors, nor a shift in γ, irrespective of the inclusion or omission of z_t, as shown by Hendry and Mizon (2012b), nor using the wrong lag value, such as x_{t-2}. Of course, none of these additional mistakes will help, but they are not the source of the problem of forecast failure.

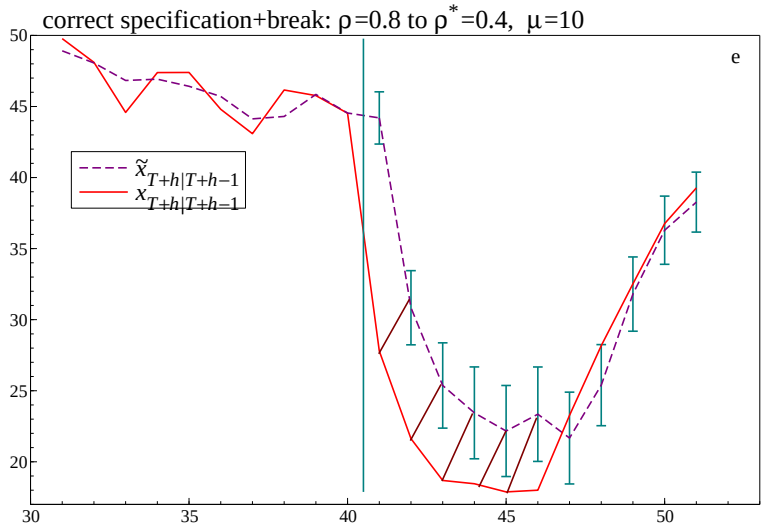
Now reconsider the same shifts in ρ but where $\mu = 10$ in:

$$x_t = \mu + \rho x_{t-1} + \gamma z_t + \epsilon_t \tag{3.27}$$

Irrespective of the correct specification of the model in-sample–whether (3.23) or (3.24) is used–there is a catastrophic impact from halving ρ, here shown in figure 3.7 panel e for forecasts from (3.23). The data are not unlike those for UK GDP, first dropping then returning. Here, the first 5 forecasts are badly biased, and each is **above the previous outcome** despite the plunging values of x, as shown by the lines connecting the previous outcomes to the next forecast. Yet forecast failure vanishes on doubling ρ again. Manifestly, the value of μ, often treated as a nuisance parameter, is actually fundamental. So is the problem just the non-zero value of the intercept–or an interaction with model mis-specification?

To clarify, again for the model which was correctly specified in-sample, consider forecasts from (3.23) for the same breaks in ρ as in (3.25) and (3.26), but setting $\mu = 0$, where that is known to the investigator. Instead, we let $\mathsf{E}[z_t] = \kappa = 10$, noting that the model is correctly specified in-sample, there is a zero intercept, and the forecasts use known future z_{T+h}. Despite all those advantageous features, forecast failure is again manifest in figure 3.7 panel f: it is almost identical to the outcome in panel e.

Figure 3.7: Correct in-sample specification with non-zero θ and changed ρ.



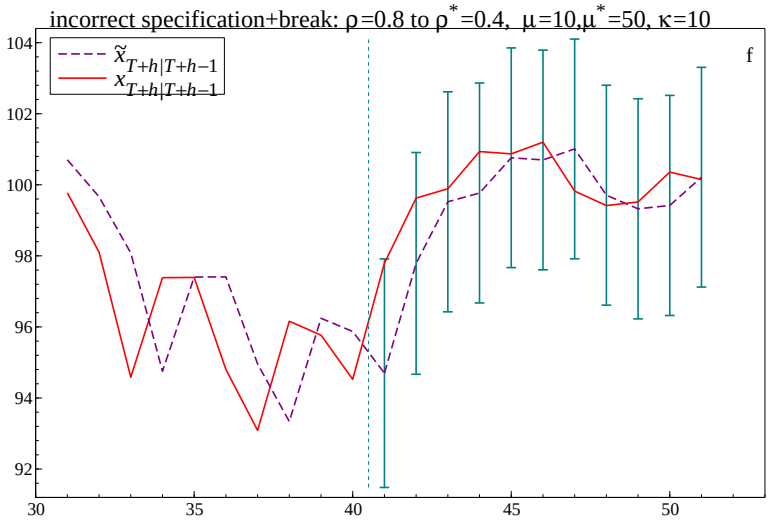
Even that is far from the whole story as can be seen as follows. When the model is incorrectly specified by omitting z_t as in (3.24), forecasting after the same breaks in ρ as in (3.25) and (3.26), but now with both $\mu = 10$, *and* $\kappa = 10$ where the former is also shifted hugely to $\mu^* = 50$ at $T = 41$ then back to $\mu = 10$ at

$T = 46$ so:

$$x_{T+h} = \mu^* + \rho^* x_{T+h-1} + \gamma z_{T+h} + \epsilon_{T+h} \tag{3.28}$$

Although almost everything seems to be wrong, there is no forecast failure using $\widehat{x}_{T+h|T+h-1} = \widehat{\mu} + \widehat{\rho} x_{T+h-1}$ over the ten forecasts in figure 3.8.

Figure 3.8: Incorrect specification with μ and ρ changed twice.



3.4.2 Understanding these forecast errors

The unique explanation for this rather strange set of outcomes is in fact simple: location shifts, or alternatively expressed in this context, shifts in the long-run mean. When they occur, forecast failure results irrespective of the goodness or otherwise of the in-sample model; when they do not occur, there is no obvious failure irrespective of model mis-specification, how many DGP parameters shift and by how much. Changes in $\mathsf{E}[x_t]$ are the culprit.

Let $\lambda = \gamma/(1-\rho)$ and rewrite the DGP in (3.21) as:

$$\Delta x_t = (\rho - 1)(x_{t-1} - \theta - \lambda(z_t - \kappa)) + \epsilon_t \tag{3.29}$$

Then from (3.21) or (3.29):

$$\mathsf{E}[x_t] = \theta = \frac{(\mu + \gamma\kappa)}{(1-\rho)} \quad \text{and} \quad \theta^* = \frac{(\mu^* + \gamma^*\kappa^*)}{(1-\rho^*)} \tag{3.30}$$

In the first three cases above, and the last, $\mathsf{E}[x_{T+h}] = 0$ before and after the shifts in ρ. In the second set of cases, $\mathsf{E}[x_{T+h}]$ shifts from $\theta = 50$ to $\theta^* = 17$ in both cases *e* and *f*, but $\theta = \theta^*$ in *g*. All models in this class are equilibrium correction so fail systematically when $\mathsf{E}[x_{T+h}]$ changes from θ to θ^*, as forecasts converge back to θ, irrespective of the new parameter values in the DGP. The class of equilibrium-correction models (EqCMS) is huge: all regressions; dynamic systems; vector autoregressions (VARs); dynamic stochastic general equilibrium systems (DSGEs); autoregressive conditional heteroskedastic processes (ARCH); and generalized ARCH (GARCH); and some other volatility models. Location shifts are a pervasive and pernicious problem affecting all EqCMs. To establish that claim formally, we next explain the taxonomy of all possible sources of forecast errors based on Hendry and Mizon (2012b).

The DGP is as in (3.21), but written as:

$$x_t = \theta + \rho(x_{t-1} - \theta) + \gamma(z_t - \kappa) + \epsilon_t \tag{3.31}$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$, $\mathsf{E}[x_t] = \theta$ and $\mathsf{E}[z_t] = \kappa$ with $\gamma \neq 0$, but z_t is omitted from the forecasting model:

$$x_t = \mu + \rho x_{t-1} + v_t \tag{3.32}$$

Here we only consider 1-step ahead forecasts facing a single break, which occurs immediately after time T, with the post-break DGP from $h = 1, \ldots$:

$$x_{T+h} = \theta^* + \rho^* \left(x_{T+h-1} - \theta^*\right) + \gamma^* \left(z_{T+h} - \kappa^*\right) + \epsilon_{T+h} \tag{3.33}$$

The estimated, mis-specified, forecasting model is denoted:

$$\widehat{x}_{T+1|T} = \widehat{\theta} + \widehat{\rho}\left(\widehat{x}_T - \widehat{\theta}\right) \tag{3.34}$$

estimated over $t = 1, \ldots, T$, with parameter estimates $(\widehat{\theta}, \widehat{\rho})$. Let $\mathsf{E}[\widehat{\theta}] = \theta_e$ and $\mathsf{E}[\widehat{\rho}] = \rho_e$. Forecasting takes place from an estimated $\widehat{x}_T$ at the forecast origin and yields the forecast error $\widehat{\epsilon}_{T+1|T} = x_{T+1} - \widehat{x}_{T+1|T}$:

$$\widehat{\epsilon}_{T+1|T} = \theta^* - \widehat{\theta} + \rho^* \left(x_T - \theta^*\right) - \widehat{\rho}\left(\widehat{x}_T - \widehat{\theta}\right) + \gamma^* \left(z_{T+1} - \kappa^*\right) + \epsilon_{T+1} \tag{3.35}$$

All the main sources of forecast error occur using (3.34) when (3.33) is the DGP:
stochastic breaks: $(\rho,\ \gamma)$ changing to $(\rho^*,\ \gamma^*)$;
deterministic breaks: $(\theta,\ \kappa)$ shifting to $(\theta^*,\ \kappa^*)$;
omitted variables: z_t excluded over both estimation and forecast periods;
biased (and inconsistent) parameter estimates: $\rho_e \neq \rho,\ \theta_e \neq \theta$ in general;
estimation uncertainty: $\mathsf{V}[\widehat{\rho},\ \widehat{\theta}] \neq \mathbf{0}$;
forecast-origin uncertainty: $\widehat{x}_T$;
innovation errors: ϵ_{T+1}.
The taxonomy of sources of forecast errors reveals all the possible effects, although we have omitted some small order interaction terms and estimation covariances of $\mathsf{O}_p(T^{-1})$ for simplicity. The calculations involved expand every term so that each final component corresponds to a single effect (e.g.):

$$\theta^* - \widehat{\theta} = (\theta^* - \theta) + (\theta - \theta_e) + \left(\theta_e - \widehat{\theta}\right) \tag{3.36}$$

so it is decomposed into shift, bias and estimation effects. Most previous taxonomies have focused on closed models, but reach similar conclusions to those discussed here.

Taxonomy for 1-step ahead forecast errors

$\widehat{\epsilon}_{T+1\|T} \simeq$	Element	Expectation	Variance
$(1-\rho^*)(\theta^*-\theta)$	(*ia*)	$(1-\rho^*)(\theta^*-\theta)$	0
$+(\rho^*-\rho)(x_T-\theta)$	(*ib*)	0	$(\rho^*-\rho)^2\,\mathsf{V}[x_T]$
$+(1-\rho)(\theta-\theta_e)$	(*iia*)	$(1-\rho)(\theta-\theta_e)$	0
$+(\rho-\rho_e)(x_T-\theta)$	(*iib*)	0	$(\rho-\rho_e)^2\,\mathsf{V}[x_T]$
$-\rho(\widehat{x}_T - x_T)$	(*iii*)	$-\rho(\mathsf{E}[\widehat{x}_T] - x_T)$	$\rho^2\mathsf{V}[\widehat{x}_T - x_T]$
$-(1-\rho)\left(\widehat{\theta}-\theta_e\right)$	(*iva*)	0	$\mathsf{O}_p(T^{-1})$
$-(\widehat{\rho}-\rho_e)(x_T-\theta)$	(*ivb*)	$\simeq 0$	$\mathsf{O}_p(T^{-1})$
$+\gamma^*(z_{T+1}-\kappa^*)$	(*v*)	0	$(\gamma^*)^2\,\mathsf{V}[z_{T+1}]$
$+\epsilon_{T+1}$	(*vi*)	0	σ_ϵ^2

(3.37)

We now consider the implications of the forecast-error taxonomy, commencing from the foot of the table:
(*vi*): the innovation error has $\mathsf{E}[\epsilon_{T+1}] = 0$ and $\mathsf{V}[\epsilon_{T+1}] = \sigma_\epsilon^2$ so there is no forecast bias, and an $\mathsf{O}_p(1)$ variance, which is irreducible if $\{\epsilon_t\}$ is an innovation;

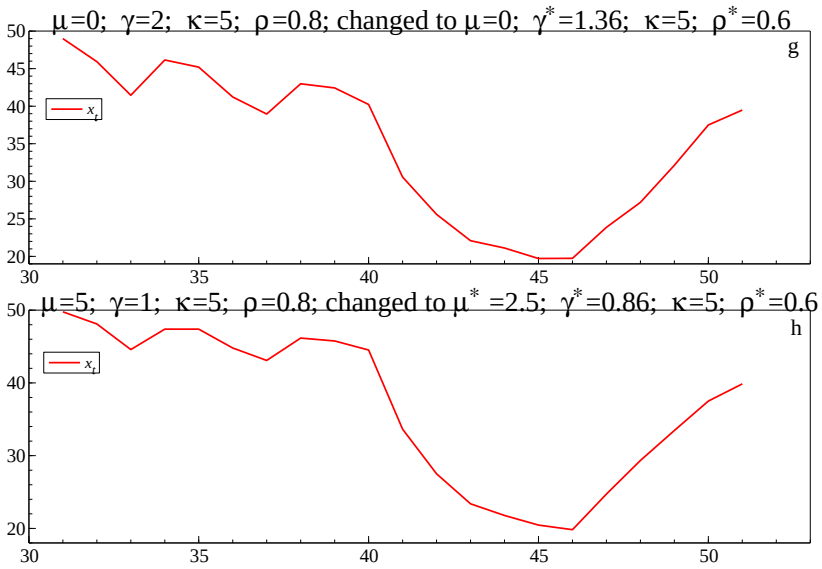
(*v*): the omitted variable leads to $\mathsf{E}[\gamma^*(z_{T+1} - \kappa^*)] = 0$ and $\mathsf{V}[\gamma^*(z_{T+1} - \kappa^*)] = (\gamma^*)^2\sigma_z^2$, so induces no bias despite its omission and the change in its parameter values, but adds an $\mathsf{O}_p(1)$ variance, reducible by including $\{z_t\}$ in the model, offset by the parameter change and an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*ivb*): slope estimation has $\mathsf{E}[(\widehat{\rho} - \rho_e)(x_T - \theta)] \simeq 0$ as $\mathsf{E}[\widehat{\rho} - \rho_e] = 0$ by definition, and also $\mathsf{E}[x_T - \theta] = 0$, but adds an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*iva*): equilibrium-mean estimation has $\mathsf{E}[(1 - \rho)(\widehat{\theta} - \theta_e)] = 0$, with an estimation variance of $\mathsf{O}_p(T^{-1})$;
(*iii*): forecast-origin uncertainty, $\mathsf{E}[\rho(\widehat{x}_T - x_T)]$, is zero only if the forecast origin is unbiasedly estimated, and will have a variance $\mathsf{O}_p(1)$ unless known for certain, emphasizing the need for good nowcasts to reduce this bias component;
(*iib*) slope mis-specification has $\mathsf{E}[(\rho - \rho_e)(x_T - \theta)] = 0$, and adds an $\mathsf{O}_p(1)$ variance unconditionally;
(*iia*) equilibrium-mean mis-specification, where $\theta \neq \theta_e$, is possible if some in-sample location shifts were not modeled, but can be eliminated by a congruent model-selection procedure that removes breaks;
(*ib*) the slope change, $\mathsf{E}[(\rho^* - \rho)(x_T - \theta)] = 0$ as $\mathsf{E}[x_T - \theta] = 0$ irrespective of $\rho^* \neq \rho$;
(*ia*) that leaves just the equilibrium-mean change–which is the fundamental problem: $\theta^* \neq \theta$ induces systematic forecast failure.

Once in-sample breaks have been removed, even from good forecast origin estimates:

$$\mathsf{E}[\widehat{\epsilon}_{T+1|T}] \simeq (1 - \rho^*)(\theta^* - \theta) \tag{3.38}$$

and that bias persists at $\widehat{\epsilon}_{T+2|T+1}$ etc., so long as (3.34) is used, even though no further breaks ensue. Keeping μ constant, but shifting ρ to ρ^*, induces a shift in θ to θ^*. The power of that insight is exemplified by:
(a) changing *both* μ and ρ by large magnitudes but with $\theta = \theta^*$ generates an outcome that is isomorphic to $\mu = \mu^* = 0$, so no break is detected as shown; and
(b) when $\mu = \mu^* = 0$, then $\kappa \neq 0$ induces forecast failure when ρ changes by shifting θ even when z_{t-1} is correctly included and is known over the forecast horizon;
(c) as figure 3.9 shows, it is feasible to essentially replicate the same break by changing μ, γ and ρ in many combinations, such that no economists or economic agents could tell what had shifted the outcome till long afterwards.

Figure 3.9: Many parameters shift.



Such results apply to all equilibrium-correction models–they fail systematically when $\mathsf{E}[x]$ changes as the model's forecasts converge to θ irrespective of value of θ^*. However, when ρ^* is changed back to ρ, the parameter values revert to those of the original DGP, so the old equilibrium is restored, and forecasts rapidly converge back to $\mathsf{E}[x]$. This suggests the original model also 'recovers' when the DGP reverts.

3.4.3 Robust forecasting devices

Robust forecasting devices may forecast better after unanticipated location shifts than any structural model in such processes, as measured by root mean-square forecast errors (RMSFEs). Clements and Hendry (1998) and Clements and Hendry (1999) provide a framework for analyzing the properties of forecasting models in wide-sense non-stationary processes, when the device being used does not coincide with the generating mechanism of the process. Clements and Hendry (2008) offer a non-technical explanation. To summarize, they explain:
(a) the recurrent episodes of systematic mis-forecasting that have occurred historically;
(b) why models that are well-specified *ex ante* need not forecast better than badly specified; since
(c) causally-relevant variables need not improve forecasting over irrelevant variables;
(d) the benefits of many of the empirical practices of forecasters, such as intercept corrections;
(e) why pooling across a range of methods and models can be beneficial as in Hendry and Clements (2004), but need not be unless carefully undertaken: see Hendry and Reade (2008b);
(f) why forecasting devices that are robust to location shifts do not experience systematic forecast failure; and so
(g) dominate in forecasting competitions, such as Makridakis and Hibon (2000): see Fildes and Ord (2002) and Clements and Hendry (2001a); hence
(h) why so-called 'naive devices' (simple adaptive devices like damped trend, differenced models, and exponentially weighted moving averages–EWMAs) can outperform; whereas
(i) equilibrium-correction models can experience systematic forecast failure: see Hendry (2006).
The simplest robust forecasting device in Hendry (2006) is discussed in Hendry and Mizon (2012b) and Castle, Fawcett, and Hendry (2011). Difference the mis-specified estimated model (3.34), so that $\Delta\widetilde{x}_{T+h|T+h-1} = \widehat{\rho}\Delta x_{T+h-1}$ or:

$$\widetilde{x}_{T+h|T+h-1} = x_{T+h-1} + \widehat{\rho}\Delta x_{T+h-1} \tag{3.39}$$

Despite using the 'wrong' $\widehat{\rho}$ for the first 5 forecasts, being incorrectly differenced, and omitting the relevant variable, nevertheless Figure 3.10 demonstrates that the robust forecasting device (3.39) avoids most of the last nine forecast errors for the DGP compared to (say) Panel *c*. The overall RMSFE is 6.6 for the in-sample model versus 5.5 for the robust forecasts here; but 3.8 versus 2.0 over the last nine forecasts, so is nearly halved, avoiding systematic failure.

Despite its non-specific assumptions, therefore, a theory of forecasting which allows for unanticipated structural breaks in an evolving economic mechanism for which the forecasting model is mis-specified in unknown ways can provide a useful basis for interpreting, and potentially circumventing, systematic forecast failure in economics. These implications seem to carry over to nowcasting.

A taxonomy of sources of forecast error for (3.39) clarifies why. Let the forecast error from the robust device be $\widetilde{\epsilon}_{T+h|T+h-1} = x_{T+h} - \widetilde{x}_{T+h|T+h-1}$. We will use accurate $\widehat{x}_T$, with $\widehat{\rho} = \rho$ for simplicity but neither is crucial. Then from (3.39):

$$\widetilde{\epsilon}_{T+1|T} = (1-\rho^*)(\theta^* - \theta) - (1-\rho^*)(x_T - \theta) + \gamma^*(z_{T+1} - \kappa^*) - \rho\Delta x_T + \epsilon_{T+1}$$

so taking expectations, using $\mathsf{E}[x_T] = \theta$, then for $h \geq 1$:

$$\mathsf{E}[x_{T+h}] \simeq \theta^* + (\rho^*)^h(\theta - \theta^*) \tag{3.40}$$

so:

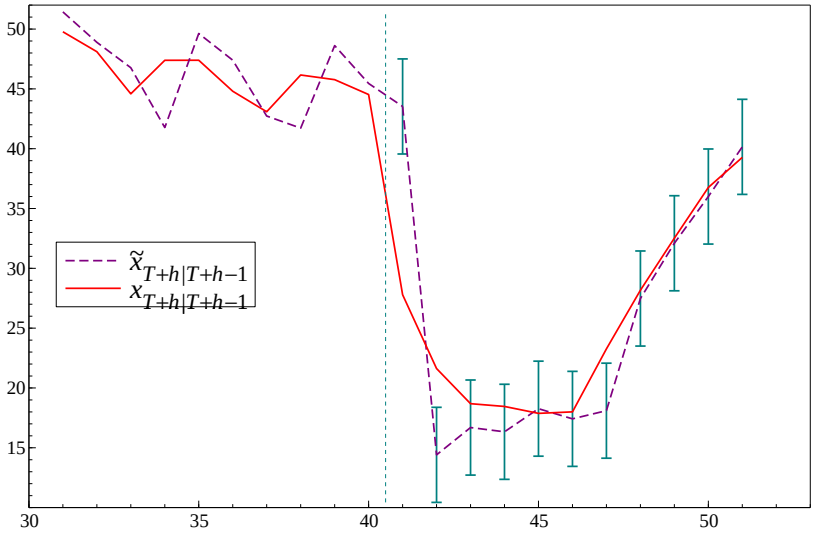
$$\mathsf{E}[\Delta x_{T+h}] \simeq (1-\rho^*)(\rho^*)^{h-1}(\theta^* - \theta) \tag{3.41}$$

hence:

$$\mathsf{E}[\widetilde{\epsilon}_{T+1|T}] \simeq (1-\rho^*)(\theta^* - \theta) \tag{3.42}$$

which is equal to the in-sample DGP forecast bias.

Figure 3.10: Differenced wrong model with changed ρ.



However, at $T+2$ the differenced-device taxonomy becomes:

$$\widetilde{\epsilon}_{T+2|T+1} = -\left(1-\rho^*\right)\left(x_{T+1}-\theta^*\right)+\gamma^*\left(z_{T+2}-\kappa^*\right)-\rho\Delta x_{T+1}+\epsilon_{T+2} \tag{3.43}$$

so from (3.40) and (3.41):

$$\mathsf{E}[\widetilde{\epsilon}_{T+2|T+1}] \simeq \left(1-\rho^*\right)\left(\rho^*-\rho\right)\left(\theta^*-\theta\right) \tag{3.44}$$

Unless ρ^* has the opposite sign to ρ, there is a valuable offset from the $-\rho\Delta x_{T+1}$ component, helping explain the forecast outcomes illustrated above even though $\rho^* \neq \rho$.

Finally, at $T+3$:

$$\mathsf{E}[\widetilde{\epsilon}_{T+3|T+2}] = \rho^*\left(\rho^*-\rho\right)\left(1-\rho^*\right)\left(\theta^*-\theta\right)$$

which is close to zero, as $\rho^*\left(1-\rho^*\right) < 0.25$. Consequently, once $h-1>2$-periods after the break, using:

$$\widetilde{x}_{T+h|T+h-1} = x_{T+h-1}+\widehat{\rho}\Delta x_{T+h-1} \tag{3.45}$$

then $\widetilde{\epsilon}_{T+h|T+h-1} = \left(1-\widehat{\rho}\right)\Delta x_{T+h-1}$ so:

$$\widetilde{\epsilon}_{T+h|T+h-1} = \left(1-\widehat{\rho}\right)\left(\rho^*-1\right)\left(x_{T+h-2}-\theta^*-\lambda^*\left(z_{T+h-1}-\kappa^*\right)\right)+\left(1-\widehat{\rho}\right)\epsilon_{T+h-1} \tag{3.46}$$

as from equation (3.29):

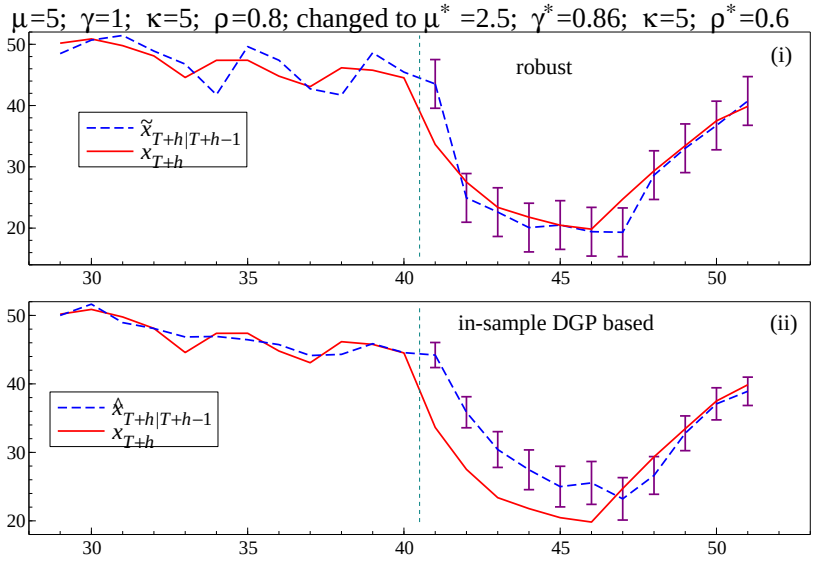
$$\Delta x_{T+h-1} = \left(\rho^*-1\right)\left(x_{T+h-2}-\theta^*-\lambda^*\left(z_{T+h-1}-\kappa^*\right)\right)+\epsilon_{T+h-1} \tag{3.47}$$

which therefore:
a] correctly reflects the new equilibrium through $\left(x_{T+h-2}-\theta^*-\lambda^*\left(z_{T+h-1}-\kappa^*\right)\right)$;
b] includes the effect from z_{T+h-1} even though that variable was omitted from the forecasting device;
c] adjusts at the speed $\left(\rho^*-1\right)$;
d] uses the well-determined in-sample estimate $\widehat{\rho}$, albeit that has shifted;
e] where $\left(1-\widehat{\rho}\right)$ in (3.46) acts like a 'damped trend'.
Consequently, $\widetilde{x}_{T+h|T+h-1}$ uses: stochastic and deterministic post-break parameters; captures all unknown omitted variables; with consistent parameters and no estimation uncertainty; and no past data measurement errors. The main drawbacks are using an additional lag; being affected by the forecast-origin error $v_{T+h|T+h-1}$ (as does (3.35) of course); and retaining the past innovation ϵ_{T+h-1}. By way of contrast, $\widehat{x}_{T+h|T+h-1}$ based on the in-sample structural model suffers from breaks in both the stochastic and deterministic parameters;

has an unknown number of omitted variables of unknown importance; with inconsistent parameter estimates and estimation uncertainty; contaminated by past data measurement errors. There are two major implications: first, there is no necessary connection between the in-sample verisimilitude of a model and its later forecasting performance; secondly, the conditional expectations at T (e.g., using the in-sample DGP) need not be the minimum RMSFE device for the outcome at $T+h$.

These taxonomy findings match previous graphs, and are little affected by estimating ρ. For the 'all parameters change' DGP, the robust device avoids almost all but the first forecast error in figure 3.11 panel (i), despite nearly all the parameters shifting, in stark contrast to the in-sample DGP forecasts in panel (ii), which show massive forecast failure for the first six forecast errors. The robust device avoids systematic forecast failure after a location shift, at an insurance cost when no shifts occur. Returning to the first correct specification and compared to Figure 3.5 from (3.22), the RMSFE of (3.39) applied to that setting is 1.0 against the estimated DGP forecasts of 0.87. Castle, Clements, and Hendry (2015) propose a class of smoothed robust devices that reduce RMSFEs in quiet periods at the cost of adjusting more slowly after shifts.

Figure 3.11: Differenced wrong model with many changed parameters.



3.4.4 Extending robust forecasting devices to I(1)

If the DGP produced non-trending I(1) data with an I(0) cointegrating relation given by $x_{t-1} - \theta - \lambda (z_t - \kappa)$, then:

$$\Delta x_t = (\rho - 1)(x_{t-1} - \theta - \lambda (z_t - \kappa)) + \epsilon_t \tag{3.48}$$

so merely rewriting the ADL would cover this additional non-stationarity. A drift term for a trend would add an intercept to (3.48), but otherwise, $\mathsf{E}[\Delta x_t] = 0$. Moreover, after a location shift, the robust forecast $\Delta \widetilde{x}_{T+h|T+h-1} = \Delta x_{T+h-1}$ is:

$$\Delta \widetilde{x}_{T+h|T+h-1} = (\rho^* - 1)(x_{T+h-2} - \theta^* - \lambda^* (z_{T+h-1} - \kappa^*)) + \epsilon_{T+h-1}$$

Thus, for $h \geq 2$, $\mathsf{E}[\Delta \widetilde{x}_{T+h|T+h-1}] \simeq 0$ as well.

3.4.5 Three empirical applications of forecasting

We will first illustrate the previous analysis by *ex post* forecasting the year on year change in Japanese exports, denoted y_t, over 2008(7)–2011(6), when they fell by more than 70%, then consider UK GDP over

two recessions. In each case, we simply contrast an autoregressive model, often difficult to outperform in forecasting, with the robust variant of the same equation obtained by differencing.

For Japanese exports, the autoregressive model selected at 1% by *Autometrics* over 2000(1)–2008(6) was (denoted by $\widehat{y}$):

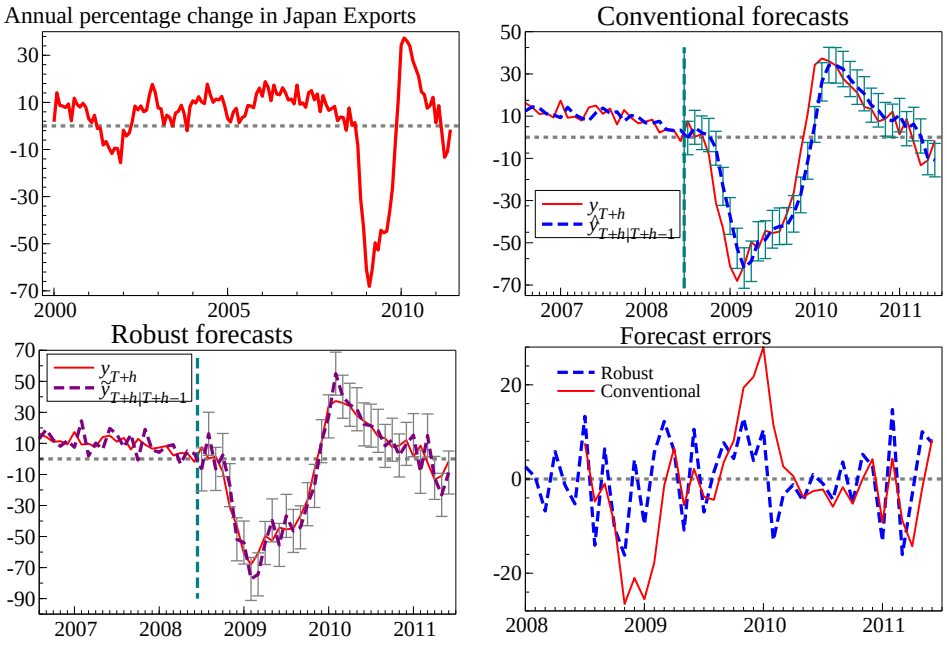
$$\begin{aligned}\widehat{y}_t &= \underset{(0.091)}{0.68}\, y_{t-1} + \underset{(0.090)}{0.26}\, y_{t-3} + \underset{(0.04)}{0.12}\, 1_{2000(2)} + \underset{(0.04)}{0.12}\, 1_{2002(1)} \\ \widehat{\sigma} &= 0.039 \quad \chi^2_{\mathsf{nd}}(2) = 0.65 \quad \mathsf{F}_{\mathsf{het}}(4,95) = 1.02 \quad \mathsf{F}_{\mathsf{ar}}(6,92) = 0.97 \quad \mathsf{F}_{\mathsf{reset}}(2,96) = 2.1\end{aligned}$$

where $1_{200z(x)}$ denote indicators for $200z(x)$. The model selected is almost 'robust' by having a near second unit root. Here, $\mathsf{F}_{\mathsf{name}}$ denotes an approximate F-test, where F_{ar} is a Lagrange-multiplier test for autocorrelation of order k: see Godfrey (1978); $\mathsf{F}_{\mathsf{het}}$ tests heteroskedasticity: see White (1980); $\mathsf{F}_{\mathsf{reset}}$ tests for non-linearity: see Ramsey (1969); and $\chi^2_{\mathsf{nd}}(2)$ is the normality test in Doornik and Hansen (2008).
The corresponding robust device was that equation differenced (denoted by $\widetilde{y}$):

$$\widetilde{y}_t = y_{t-1} + 0.95\Delta y_{t-1} \qquad \widetilde{\sigma} = 0.0755$$

Their respective forecasts are shown in Figure 3.12.

Figure 3.12: Forecasts of year-on-year changes in Japan's exports, 2008(7)–2011(6).



This is a typical pattern: the robust device avoids forecast failure, at an insurance cost when no shifts occur. Overall, their comparative RMSFEs are $\mathbf{0.124}$ versus $\mathbf{0.098}$ as highlighted by Figure 3.13.

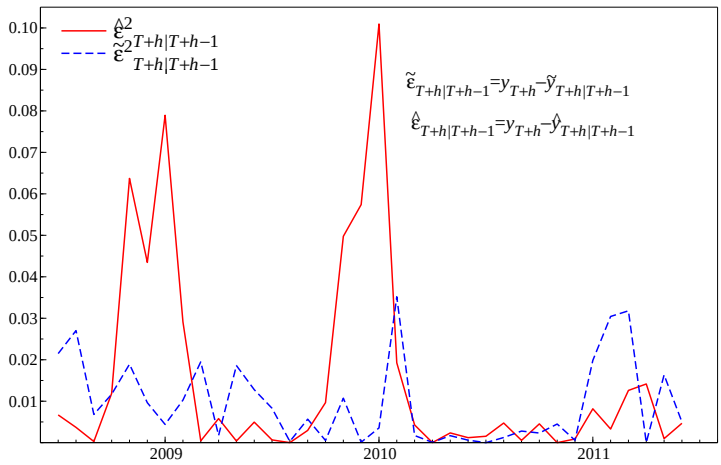
Forecasting UK GDP

The autoregressive model selected at 1% by *Autometrics* over 1989(2)–2007(4) was:

$$\begin{aligned}\widehat{y}_t &= \underset{(0.083)}{0.537}\, y_{t-1} + \underset{(0.001)}{0.003} - \underset{(0.003)}{0.018}\, 1_{1990(3)} \\ \widehat{\sigma} &= 0.0032 \quad \chi^2_{\mathsf{nd}}(2) = 2.27 \quad \mathsf{F}_{\mathsf{ar}}(5,6) = 1.60 \\ \mathsf{R}^2 &= 0.51 \quad \mathsf{F}_{\mathsf{het}}(2,71) = 0.05 \quad \mathsf{F}_{\mathsf{reset}}(2,70) = 1.43\end{aligned}$$

where $1_{1990(3)}$ is an indicator for $1990(3)$. The corresponding robust device was the difference:

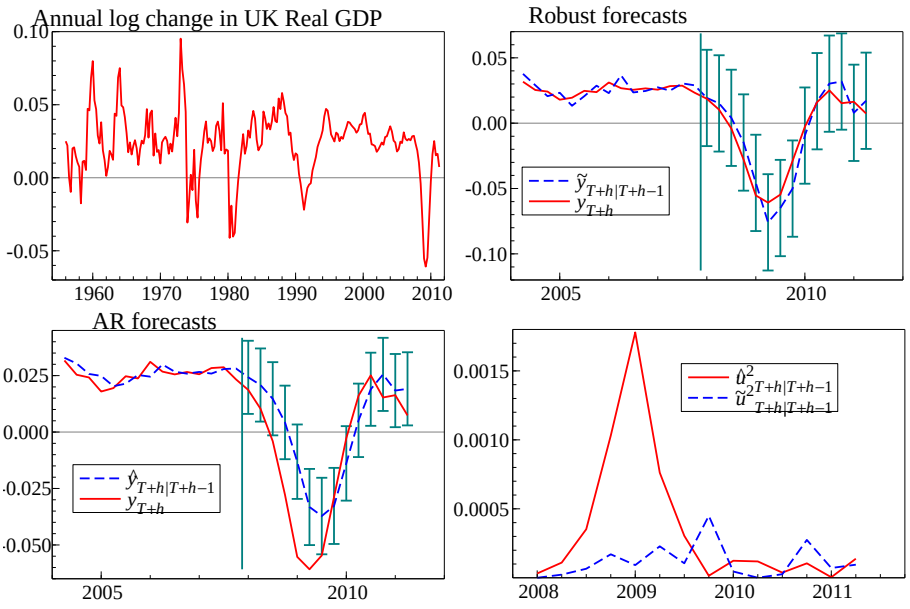
Figure 3.13: Squared forecast-error comparison.



$$\widetilde{y}_t = y_{t-1} + 0.5\Delta y_{t-1} \qquad \widetilde{\sigma} = 0.0053$$

Their respective forecasts across 2008–2011 are in Figure 3.14, and the RMSFEs are now 0.07 versus 0.04, so differencing nearly halves it, as panel d reveals.

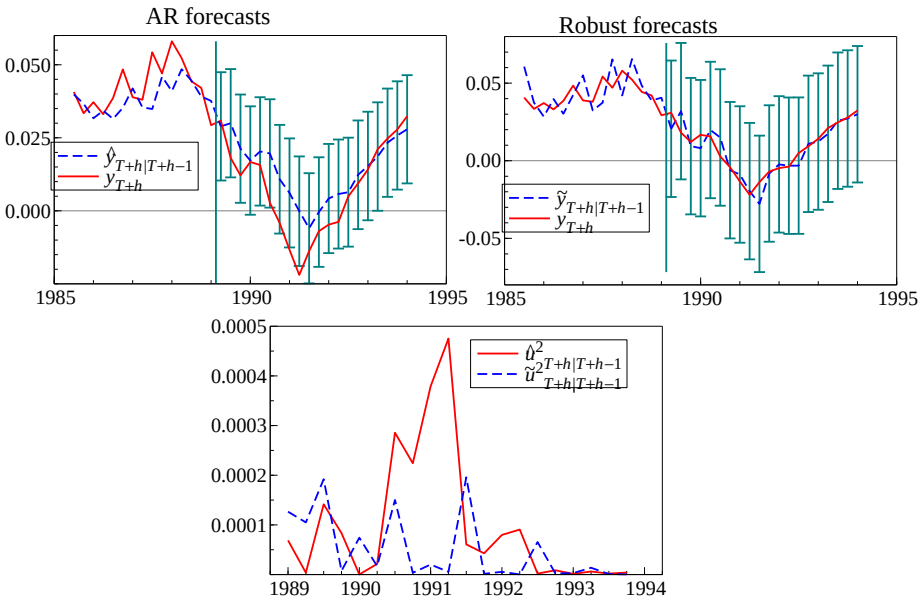
Figure 3.14: Data, forecasts and squared forecast errors over 2008–2011.



Forecasts over an earlier debacle

HM Treasury also badly mis-forecast the UK recession in the early 1990s, so we reconsider that debacle in Figure 3.15. The forecasts from an autoregression are again outperformed by the robust differenced device. Note that an agency can wait until forecast failure has occurred before switching to a robust method, as the latter is no better at forecasting shifts: Castle, Fawcett, and Hendry (2010) and Castle, Fawcett, and Hendry (2011) discuss the possibilities of forecasting location shifts.

Figure 3.15: Forecast failure in the early 1990s.



3.5 Automatic model selection

In this section, we describe the basis of automatic model selection when there are more variables than observations. Automatic model selection for nowcasting all disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts can be undertaken following the approach described in Castle, Doornik, and Hendry (2012) and Hendry and Doornik (2014) based on *Autometrics* with impulse-indicator saturation: see Doornik (2009a) on the former, and Hendry et al. (2008) and Johansen and Nielsen (2009) for the latter. The theory of reduction explains the existence of a local data generation process (LDGP) for any subset of variables, and the relationship of any model to that LDGP: see e.g. Hendry (2009) for a recent exposition. The idea behind general-to-specific (Gets) model selection is to locate a good approximation to that LDGP, characterized by its being a congruent representation and encompassing the evidence on the LDGP both directly and via information embodied in other models. To do so, one commences from the most general model that nests all the candidate variables, their lags, functional form transformations and possible breaks, intrinsically leading to having more variables, N, than observations T. Castle, Doornik, and Hendry (2013) consider using all the variables jointly with their principal components, extended to forecasting by Castle, Clements, and Hendry (2013).

To model complex equations, we rely on general to specific modeling approaches: see Campos et al. (2005). The unknown DGP generates the data, but empirical modeling will always deal with a subset of variables in that DGP, creating the LDGP as the generating process in the space of the variables under analysis: see Hendry (2009). The approach, therefore, aims to construct a set of data based on broad theoretical assumptions, which nests the LDGP, then within this set, reduce the model from its general form down to a specific well-specified and undominated representation. This is a two step procedure. One: define a set of N variables that should include the LDGP as a sub-model. Two: starting with that general model as a good approximation of the overall properties of the data, reduce its complexity by removing insignificant variables, while checking that at each reduction the validity of the model is preserved. This is the basic framework of *Gets* modeling.

This section introduces theoretical concepts of model selection, their use in mis-specification testing, followed by the introduction of impulse-indicator saturation (IIS) and its generalized version. All these concepts are then united and applied through the automatic search algorithm *Autometrics*. The algorithm combines these features through automated selection based on *Gets* while handling more variables than observations with IIS

for detecting breaks and outliers, and mis-specification testing.

The theory of reduction characterizes the operations implicitly applied to the DGP to obtain the local LDGP. Choosing to analyze a set of variables, denoted $\mathbf{y}_t, \mathbf{x}_{1,t}$, determines the properties of the LDGP, and hence of any models of $\mathbf{y}_t$ given $\mathbf{x}_{1,t}$ (with appropriate lags, non-linear transforms thereof, etc.). A congruent model is one that matches the empirical characteristics of the associated LDGP, evaluated by a range of mis-specification tests: see e.g., Hendry (1995) and section 3.5.2. A model is undominated if it encompasses, but is not encompassed by, all other sub-models: see e.g., Mizon and Richard (1986), Hendry and Richard (1989) and Bontemps and Mizon (2008).

3.5.1 Autometrics

Autometrics is the latest installment in the automated *Gets* methodology and is available in the *OxMetrics* software package: see Doornik (2009a). The algorithm is based on the following main components:[6]

1. GUM: The general unrestricted model (GUM) is the starting point of the search. The GUM should be specified based on broad theoretical considerations to nest the LGDP.
2. Pre-Search: prior to specific selection, a pre-search lag reduction can be implemented to remove insignificant lags, speeding up selection procedures and reducing the fraction of irrelevant variables selected (denoted the gauge of the selection process). Pre-search is only applied if the number of variables does not exceed the number of observations ($N < T$)
3. Search Paths: *Autometrics* uses a tree search to explore paths. Starting from the GUM, *Autometrics* removes the least significant variable as determined by the lowest absolute t-ratio. Each removal constitutes one branch of the tree. For every reduction, there is a unique sub-tree which is then followed; each removal is back-tested against the initial GUM using an F-test. If back-testing fails, no sub-nodes of this branch are considered (though different variants of this removal exist). Branches are followed until no further variable can be removed at the pre-specified level of significance α. If no further variable can be removed, the model is considered to be terminal.
4. Diagnostic Testing: each terminal model is subjected to a range of diagnostic tests based on a separately chosen level of significance. These tests include tests for normality (based on skewness and kurtosis), heteroskedasticity (for constant variance using squares), the Chow (1960) test (for parameter constancy in different samples), and residual autocorrelation and autoregressive conditional heteroskedasticity. Parsimonious encompassing of the feasible general model by sub-models both ensures no significant loss of information during reductions, and maintains the null retention frequency of *Autometrics* close to α: see Doornik (2008). Both congruence and encompassing are checked by *Autometrics* when each terminal model is reached after path searches, and it backtracks to find a valid less reduced earlier model on that path if any test fails. This repeated re-use of the original mis-specification tests as diagnostic checks on the validity of reductions does not affect their distributions: see Hendry and Krolzig (2003).
5. Tiebreaker: as a result of the tree search, multiple valid terminal models can be found. The union of these terminal models is referred to as the terminal GUM. As a tiebreaker to select a unique model, the likelihood-based Schwarz (1978) information criterion (SIC) is used, though other methods are also applicable, and terminal models should be considered individually.

In simulation experiments, models are primarily evaluated based on three concepts: gauge, potency and the magnitudes of the estimated parameters' RMSEs around the DGP values: see Doornik and Hendry (2013). Gauge describes the retention of irrelevant variables when selecting (i.e., variables that are selected but do not feature in the DGP). Potency measures the average retention frequency of relevant variables (variables that are selected and feature in the DGP). Low gauge (close to zero) and high potency (close to 1) are preferred, as are small RMSEs.

[6] The following explanation is based on the exposition in Hendry and Pretis (2013).

The main calibration decision in the search algorithm is the choice of significance level α at which selection occurs. Selection continues until retained variables are significant at α, though it can be the case that variables in the final model are also retained at a level above α if removal leads to diagnostic tests failing. α is approximately equal to the gauge of selection. Further, the choice of diagnostic tests and lag length selection for residual autocorrelation and autoregressive conditional heteroskedasticity need to be set.

3.5.2 Mis-specification testing

Using a large number of variables with IIS (discussed in the next section) also provides a new view of model evaluation: to avoid mis-specification and non-constancy, start as general as possible within the theoretical framework, using all the available data unconstrained by $N > T$, (where N is the number of variables and T the number of observations) retaining the theory inspired variables and only selecting over the additional candidates.

Even so, this approach does not obviate the need to test the specification of the auxiliary hypotheses against the possibility that the errors are not independent, are heteroskedastic (non-constant variance) or non-normal, that the parameters are not constant, that there is unmodeled non-linearity, and that the conditioning variables are not independent of the errors. When $N << T$, the first five are easily tested in the initial general model; otherwise their validity can be checked only after a reduction to a feasible sub-model. Congruence is essential not only to ensure a well specified final selection, but also for correctly-calibrated decisions during selection based on Gaussian significance levels, which IIS will help ensure.

The model selection approach introduced here allows for more variables than observations to be used in modeling ($N > T$). For *Autometrics* this was first introduced through impulse-indicator saturation, which we now discuss, and has recently been extended to the general case.

3.5.3 Impulse-indicator saturation

Autometrics handles the $N > T$ problem by a mixture of expanding and contracting searches that seek all the variables relevant at the chosen significance level α, set such that αN remains small (e.g., unity). Multiple breaks are tackled conjointly by impulse-indicator saturation (IIS) which adds an impulse indicator for every observation. Hendry et al. (2008) establish one feasible algorithm, and derive the null distribution for an IID process, and Johansen and Nielsen (2009) generalize their findings to general dynamic regression models (possibly with unit roots), and show that there is a very small efficiency loss under the null of no breaks when αT is small, despite investigating the potential relevance of T additional variables. Castle, Doornik, and Hendry (2012) examine the ability of IIS to detect multiple breaks, and show it can find up to 20 breaks in 100 observations.

The numbers, timings and magnitudes of breaks in models are usually unknown, and are obviously unknown for unknowingly omitted variables, so a 'portmanteau' approach is required that can detect location shifts at any point in the sample while also selecting over many candidate variables. To check the null of no outliers or location shifts in a model, impulse-indicator saturation (IIS) creates a complete set of indicator variables $\left\{1_{\{j=t\}}\right\}$ equal to unity when $j = t$ and equal to zero otherwise for $j = 1, \ldots, T$, and includes these in the set of candidate regressors. Although this creates T variables when there are T observations, in the 'split-half' approach analyzed in Hendry et al. (2008), a regression first includes $T/2$ indicators. By dummying out the first half of the observations, estimates are based on the remaining data, so any observations in the first half that are discrepant will result in significant indicators. The location of the significant indicators is recorded, then the first $T/2$ indicators are replaced by the second $T/2$, and the procedure repeated. The two sets of significant indicators are finally added to the model for selection of the significant indicators. The distributional properties of IIS under the null are analyzed in Hendry et al. (2008), and extended by Johansen and Nielsen (2009) to both stationary and unit-root autoregressions.

Figure 3.16: 'Split-sample' IIS search under null.

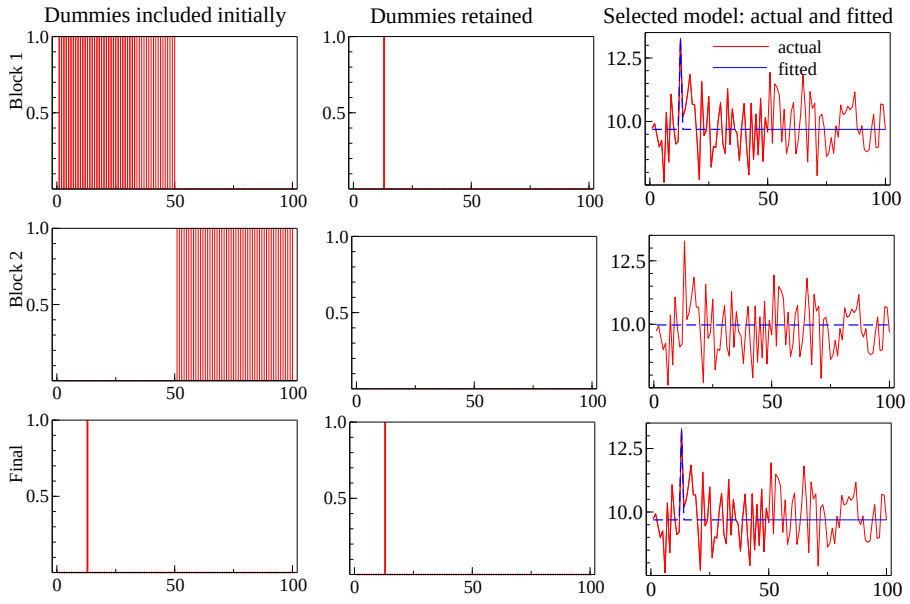


Figure 3.16 illustrates the 'split-half' approach for $y_t \sim \mathsf{IN}\left[\mu, \sigma_y^2\right]$ for $T = 100$ selecting indicators at a 1% significance level (denoted α).[7] The three rows correspond to the three stages: the first half of the indicators, the second half, then the selected indicators combined. The three columns respectively report the indicators entered, the indicators finally retained in that model, and the fitted and actual values of the selected model. Initially, although many indicators are added, only one is retained. When those indicators are dropped and the second half entered (row 2), none is retained. Now the combined retained dummies are entered (here just one), and selection again retains it. Since $\alpha T = 1$, that is the average false null retention rate.

We next illustrate IIS for a location shift of magnitude λ over the last k observations:

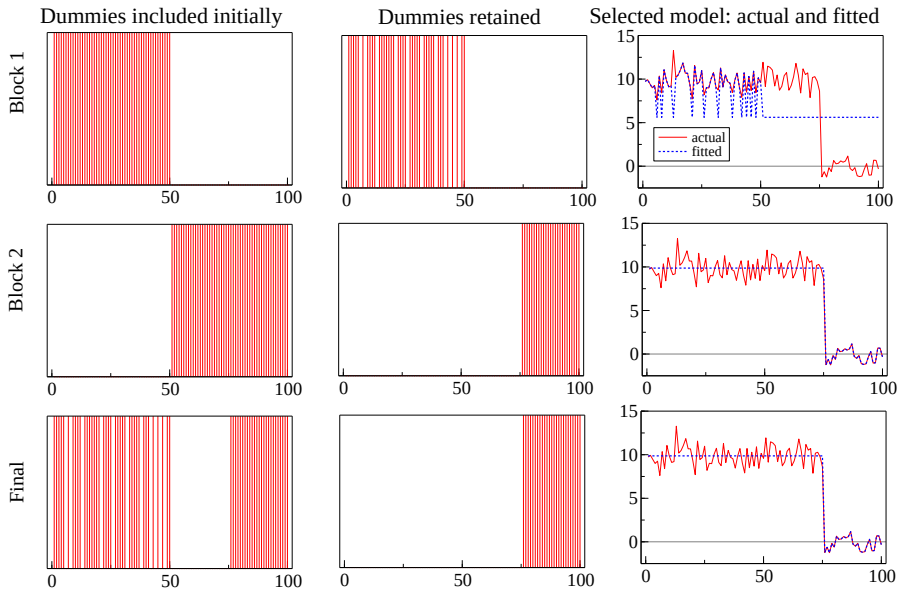
$$y_t = \mu + \lambda 1_{\{t \geq T-k+1\}} + \varepsilon_t \tag{3.49}$$

where $\varepsilon_t \sim \mathsf{IN}\left[0, \sigma_\varepsilon^2\right]$ and $\lambda \neq 0$. The optimal test in this setting would be a t-test for a break in (3.49) at $T - k + 1$ onwards, but requires precise knowledge of the location-shift timing, as well as knowing that it is the only break and is the same magnitude break thereafter. Figure 3.17 records the behavior of IIS for a mean shift in (3.49) of $10\sigma_\varepsilon$ occurring at $0.75T = 75$. Initially, many indicators are retained (top row), as there is a considerable discrepancy between the first-half and second-half means. When those indicators are dropped and the second set entered, all those for the period after the location shift are now retained. Once the combined set is entered (despite the large number of dummies) selection again reverts to just those for the break period. Under the null of no outliers or breaks, any indicator that is significant on a sub-sample would remain so overall, but for many alternatives, sub-sample significance can be transient, due to an unmodeled feature that occurs elsewhere in the data set. Thus there is an important difference between 'outlier detection' procedures and IIS.

While IIS is perhaps surprising initially, many well-known statistical procedures are variants of IIS. The Chow (1960) test corresponds to sub-sample IIS over $T - k + 1$ to T, but without selection, as Salkever (1976) showed, for testing parameter constancy using indicators. Recursive estimation is equivalent to using IIS over the future sample, and reducing the indicators one at a time. Johansen and Nielsen (2009) relate IIS to robust estimation, and show that under the null of no breaks, outliers or data contamination, the average cost of applying IIS is the loss of αT observations. Thus, at $\alpha = 0.01$, for $T = 100$ one observation is 'dummied out' by chance despite including 100 'irrelevant' impulse indicators in the search set and checking for location

[7] Figures 3.16 and 3.17 are also reported in Hendry and Pretis (2013) and Hendry and Doornik (2014).

Figure 3.17: 'Split-sample' IIS search under the alternative.



shifts and outliers at every data point. Retention of theory variables is feasible during selection with IIS as in Hendry and Johansen (2015), as is jointly selecting over the non-dummy variables, and IIS can be generalized to multiple splits of unequal size. While IIS entails more candidate variables than observations as $N+T > T$, selection is feasible as *Autometrics* undertakes expanding as well as contracting block searches (see next section). Non-linear model selection (including threshold models) is examined in Castle and Hendry (2011) and Castle and Hendry (2014b).

For a single location shift, Hendry and Santos (2010) show the detection power is determined by the magnitude of the shift τ; the length of the break interval $T - T^*$, which determines how many indicators need to be found; the error variance of the equation σ_η^2; and the significance level, α, where a normal-distribution critical value c_α, is used by the IIS selection algorithm. Castle, Doornik, and Hendry (2012) establish the ability of IIS to detect multiple location shifts and outliers, including breaks close to the start and end of the sample, as well as correcting for non-normality.

3.5.4 General case of $N > T$

The idea of generalizing using more variables than observations from IIS to all forms of independent variables is introduced by Hendry and Krolzig (2005) and implemented by Doornik (2007). Suppose there are $N = \sum_{j=1}^{N} n_j$ total regressors such that $N > T$ and $n_j < T$ for all j. Consequently the total number of variables N exceeds the number of observations T but total variables can be partitioned into blocks n_j each smaller than T. Their approach suggests randomly partitioning the set of variables into blocks of n_j, apply *Gets* to each block retaining the selected variables and crossing the groups to mix variables. The next step is to use the union of selected variables from each block to form a new initial model and repeat the process until the final union of selected variables is sufficiently small. *Autometrics* implements a different variant of this algorithm to handle the general case of $N > T$ using a combination of expanding and contracting block searches that 'learns' as it proceeds.

In the general case of $N > T$ and IIS, *Autometrics* groups variables into two categories: selected and not selected: see Doornik (2007). Not currently selected variables are split into sub-blocks and the algorithm proceeds by alternating between two steps: first, the expansion step, selection is run over not-selected sub-blocks to detect omitted variables. Second, the reduction step, a new selected set is found by running selection on the system augmented with the omitted variables found in step one. This is repeated until the dimensions

of the terminal model are small enough and the algorithm converges, so the final model is unchanged by further searches for omitted variables.

Autometrics has been applied successfully in a range of fields: see, for example, Hendry and Mizon (2011a) on US food expenditure, Bårdsen, Hurn, and McHugh (2010) on unemployment in Australia, Bårdsen, den Reijer, Jonasson, and Nymoen (2012) for a macromodel of Sweden, Mariscal and Powell (2014) to commodity price booms, and Castle, Doornik, and Hendry (2011) and Hendry and Doornik (2014) for comparisons with other selection methods.

Nevertheless overall selections should be interpreted carefully. Successful identification of the underlying LGDP can be adversely affected by collinearity of the independent variables. Most simulations of *Autometrics* with large numbers of variables use orthogonal regressors, which makes selection easier. Furthermore, when $N > T$, in the block selection algorithm of *Autometrics*, adding or dropping a variable from the initial GUM may change the block partitioning of variables, so the selection is not invariant to the initial specification.

3.6 Producing 'good' nowcasts

In this section, we apply the preceeding analyses of forecasting and model selection to nowcasting an aggregate from disaggregates. The intent of Gets is to find 'good' LDGP models as defined in section 3.5, even though Clements and Hendry (1998) and Clements and Hendry (1999) show that one cannot prove that the pre-existing LDGP is the best model for forecasting.

A timely data source that provided an accurate 'measure' of a required aggregate variable is clearly preferable to seeking 'good forecasts' of it. Nevertheless, there will usually be a role for forecasting 'preliminary estimates', so good forecasting methods are likely to remain important, albeit which methods might be 'best' is an empirical issue as seen in our extensive discussion of forecasting in Section 3.4. Nowcasting is also partly a 'signal extraction' problem for missing data entering the aggregate, but if the first announcements can be systematically improved by forecasting them either directly or via the disaggregates, then the quality of the resulting aggregate is bound to be better. Three problems additional to the seven discussed in the Introduction inhibit achieving that outcome. All three considerations apply to nowcasting an aggregate either directly or via its disaggregates.

First, the objective functions of the users of a nowcaster's output are almost always unknown. A convenient approximation is to assume a quadratic loss, in which case the aim of nowcasting becomes to find a forecast $\widehat{y}_{T|T-\delta}$ of the aggregate y_T which solves:

$$\underset{\widehat{y}_{T|T-\delta}}{\arg\min}\, \mathsf{E}_{\mathsf{D}_{y_T}}\left[y_T - \widehat{y}_{T|T-\delta}\right]^2 \tag{3.50}$$

where $\widehat{y}_{T|T-\delta} = g_{T-\delta}(\widehat{\mathcal{J}}_{T-\delta})$, and $g_{T-\delta}(\cdot)$ is the relevant function of the available measured information set, $\widehat{\mathcal{J}}_{T-\delta}$, that estimates the actual information, $\mathcal{I}_{T-\delta}$, on which y_T depends via $y_T = f_T(\mathcal{I}_{T-\delta})$. We have dated information as $T-\delta$ for $\delta > 0$, since some evidence must be unavailable at T to necessitate nowcasting. To proceed, we will assume that (3.50) does indeed provide the objective function, as it is difficult to see why either direction of asymmetry should dominate.

The second difficulty is that the information to be included in $\widehat{\mathcal{J}}_{T-\delta}$ is unknown, and could comprise just the history of the relevant series (via a univariate time-series model), current and past data on other related time series, past information about revisions etc. We address this issue by proposing automatic model selection either directly for the aggregate itself, or for nowcasting the disaggregates allowing for all the available information, multiple past breaks and contemporaneous location shifts as in Castle, Doornik, and Hendry (2012), then relating the aggregate to its own past and all the disaggregates as in Hendry and Hubrich (2011). Section 3.6.1 considers relating aggregates to disaggregates using automatic model selection (which section 3.5.1 discussed).

The third, and most serious, problem is obtaining $\widehat{y}_{T|T-\delta}$. The conditional expectation $\mathsf{E}_{\mathsf{D}_{y_T}}[y_T|\widehat{\mathcal{J}}_{T-\delta}]$ of y_T given $\widehat{\mathcal{J}}_{T-\delta}$ is the solution to (3.50), namely the minimum mean-square error predictor, only when no breaks occur. Unfortunately, $\mathsf{E}_{\mathsf{D}_{y_T}}[\cdot]$ is not known at T either, since the statistical process generating $\{y_t\}$ is never known, and in economics is always wide-sense non-stationary, namely its distribution changes over time from both stochastic trends and location shifts, reflected in our notation $\mathsf{D}_{y_T}(\cdot)$. Thus, the form of the optimal predictor $\widehat{y}_{T|T-\delta}$ is never known , and $\mathsf{E}_{\mathsf{D}_{y_{T-\delta}}}[y_T|\widehat{\mathcal{J}}_{T-\delta}]$, which might be available, need not be a good forecast device when location shifts occur, as it is calculated over the wrong distribution: see e.g., Castle, Doornik, Hendry, and Nymoen (2014) and Hendry and Mizon (2014). Thus, the realized mean square forecast error (MSFE) given (3.50) is:

$$\mathsf{E}_{\mathsf{D}_{y_T}}\left[f_T\left(\mathcal{I}_{T-\delta}\right)-g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right)\right]^2 \tag{3.51}$$

leading to the following abbreviated taxonomy of forecast errors $\widehat{u}_{T|T-\delta}=y_T-\widehat{y}_{T|T-\delta}$:

$$\begin{aligned}
\widehat{u}_{T|T-\delta} &= f_T\left(\mathcal{I}_{T-\delta}\right)-g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right)\\
&= f_T\left(\mathcal{I}_{T-\delta}\right)-f_{T-\delta}\left(\mathcal{I}_{T-\delta}\right)\text{–location shift (section 3.4.3)}\\
&\quad +f_{T-\delta}\left(\mathcal{I}_{T-\delta}\right)-g_{T-\delta}\left(\mathcal{I}_{T-\delta}\right)\text{–model mis-specification (section 3.4 and 3.5.2)}\\
&\quad +g_{T-\delta}\left(\mathcal{I}_{T-\delta}\right)-g_{T-\delta}\left(\mathcal{J}_{T-\delta}\right)\text{–reduced information (section 3.6.1)}\\
&\quad +g_{T-\delta}\left(\mathcal{J}_{T-\delta}\right)-g_{T-\delta}\left(\widehat{\mathcal{J}}_{T-\delta}\right)\text{–measurement error (section 3.6.8)}
\end{aligned} \tag{3.52}$$

Every term is potentially non-zero and doubtless ever present, but the expected values of the terms from model mis-specification and reduced information (which is not information about future location shifts) are zero. Thus, location shifts (the focus above) and measurement errors (the focus of nowcasting) seem likely to be the most pernicious problems when they occur.

3.6.1 Nowcasting aggregates from disaggregates

The variable of interest, y_T, say GDP, is an aggregate variable, comprising $y_t=\sum_{i=1}^{N_T}w_iy_{i,t}$ where $y_{i,t}$ are the disaggregates and w_i are the weights, which could be changing over time, as could the number of disaggregates. Data are released with varying time delays, such that at time T some components of y_T will be observed and some will be unavailable until $T+\delta$. For $y_{i,T}$, the $i=1,\ldots,J_T$, components are known at T and the $i=J_T+1,\ldots,N_T$, components are unknown at T. The two sets are not uniquely defined, so an individual element can switch back and forth between the sets over time depending on how the information is accrued, and J_T is not fixed. Hence, a forecasting strategy needs to be flexible enough to allow for changes in the timing of releases. As some components of y_T are unknown at T, a nowcast is computed. There are three alternative methods, see Hendry and Hubrich (2011), which we first describe, then address selection and estimation in section 3.6.5:

I Forecast the aggregate using only aggregate information, some of which perforce must be lagged one period:

$$\widehat{y}_{T|T-\delta^*}=f\left(\mathbf{Y}^s_{T-1},\mathbf{Z}^s_{T-\delta^*}\right) \tag{3.53}$$

where $\mathbf{Y}^s_{T-1}=(y_{T-1},\ldots,y_s)$, and $\mathbf{Z}^s_{T-\delta^*}=(\mathbf{z}_{T-\delta^*},\ldots,\mathbf{z}_s)$ is a vector of current and lagged conditioning variables such as surveys or leading indicators, which may be more recent than the latest aggregate observation, with $\delta^*\leq\delta\leq 1$.

II Forecast those disaggregates that are unknown at T:

$$\widetilde{y}_{i,T|T-\delta^*}=f\left(g\left(\mathbf{y}^0_{J_T,T}\right),\mathbf{Y}^s_{T-1}\mathbf{Z}^s_{T-\delta^*}\right),\quad i=J_T+1,\ldots,N_T \tag{3.54}$$

where $g(\cdot)$ is a function, perhaps just a subset, of the known data in $\mathbf{y}^0_{J_T,T} = (y_{1,T} \dots y_{J_T,T})$. Then the forecasts of the $\widetilde{y}_{i,T|T-\delta^*}$ could be aggregated with the known data:

$$\widetilde{y}_{T|T} = \sum_{i=1}^{J_T} w_i y_{i,T} + \sum_{i=J_T+1}^{N_T} w_i \widetilde{y}_{i,T|T-\delta^*} \tag{3.55}$$

Typically the weights are taken as given but could also be forecast; we abstract from the issue of weights in the subsequent analysis, as they are known with GDP data. Method [II] nests method [I], and could be applied even if $J_T = 0$, as would occur within a quarter, so high-frequency updates are possible in advance of measurements arriving. Since the measured variables in J_T change over time, section 3.6.6 discusses how to specify $g(\cdot)$.

III Forecast the aggregate, conditioning on both aggregate and disaggregate information. There are two possible methods.

a. In order to apply standard forecasting procedures, a balanced panel is needed:

$$\overline{y}_{T|T-\delta} = f\left(\mathbf{Y}^s_{T-1}, \mathbf{y}^0_{J_T,T}, \mathbf{Z}^s_{T-\delta^*}\right), \quad \forall i \in N_T, \tag{3.56}$$

where the conditioning information includes information dated T, $T-\delta$ and previous, resulting in a δ-step ahead forecast.

b. Alternatively, an unbalanced panel can be used, conditioning on whatever information is also available at T:

$$\overline{\overline{y}}_{T|T} = f\left(\mathbf{Y}^s_{T-1}; \mathbf{y}^0_{J_T,T}, y_{J_T+1,T-\delta}, \dots, y_{N_T,T-\delta}; \mathbf{Z}^s_{T-\delta^*}\right),$$

resulting in a large unbalanced panel in which there are missing observations at the end of the sample: see Wallis (1986) for a discussion of the ragged-edge problem.

Approach [I] has been the most common approach in the literature, with a focus on predicting the revision process at the aggregate level, see Lee et al. (2008). However, the aggregate approach relies on a reduced information set, $\mathcal{J}_{T-1} \subset \mathcal{I}_{T-1}$, where $\mathcal{I}_{T-1}$ is the full information set including disaggregate information. As discussed in section 3.3, unpredictability is relative to the information set used, so using a subset of information will result in less accurate predictions. Although we do not use [I] here for the missing disaggregates, instead focusing on using the additional new disaggregate information, (3.53) enables *ex ante* forecasts of all disaggregates to be available before measured data arrive to be compared with those observed outcomes to test for any location shifts. When all disaggregates are observed, Espasa and Mayo-Burgos (2012) discuss how to incorporate cointegrated and cyclical information about disaggregates to improve the aggregate forecasts relative to unrestricted estimation.

We now show analytically that when interest focuses on predicting the aggregate, then nothing is lost by doing so directly from disaggregate information, without predicting the disaggregates, i.e., method [III]. $N = 2$ suffices to illustrate the analysis, which generalizes to many components:

$$y_T = w_1 y_{1,T} + (1 - w_1)\, y_{2,T} \tag{3.57}$$

with weights w_1 and $w_2 = (1 - w_1)$, where for simplicity we take the weights to be constant over time, as in a simple sum aggregate. Assume that $J = 0$, so the contemporaneous disaggregates are not observed, and information is only available from $T-1$. The DGP for the disaggregates is assumed to be:

$$y_{i,t} = \gamma_i' \mathbf{x}_{t-1} + \eta_{i,t}$$

where $\mathbf{x}_{t-1} = (y_{1,t-1}, y_{2,t-1}; \mathbf{z}_{t-1})'$ denotes all the available information, which includes the lagged disaggregates, so:

$$\mathsf{E}_{T-1}\left[y_{i,T} \mid \mathbf{x}_{T-1}\right] = \gamma_i' \mathbf{x}_{T-1} \text{ for } i = 1, 2 \tag{3.58}$$

Aggregating the two terms in (3.58), delivers:

$$\mathsf{E}_{T-1}\left[y_T \mid \mathbf{x}_{T-1}\right] = \sum_{i=1}^{2} w_i \mathsf{E}_{T-1}\left[y_{i,T} \mid \mathbf{x}_{T-1}\right] = \left(\sum_{i=1}^{2} w_i \gamma_i'\right) \mathbf{x}_{T-1} = \psi' \mathbf{x}_{T-1} \tag{3.59}$$

say. Predicting y_T directly from $\mathbf{x}_{T-1}$ yields:

$$\mathsf{E}_{T-1}\left[y_T \mid \mathbf{x}_{T-1}\right] = \pi' \mathbf{x}_{T-1} \tag{3.60}$$

Since the left-hand sides of (3.59) and (3.60) are equal:

$$\pi' \mathbf{x}_{T-1} = \psi' \mathbf{x}_{T-1} \tag{3.61}$$

so nothing is lost by predicting y_T directly, instead of aggregating component predictions once the same information set $\mathbf{x}_{T-1}$ is used for both, i.e. there is no benefit to method [IIIa] over [II].

Hendry and Hubrich (2011) show that this analysis generalizes to non-constant weights and changing parameters, and demonstrate that forecast-origin breaks equally affect combining disaggregate forecasts to forecast an aggregate as in (3.59) or forecasting the aggregate directly as in (3.60), so there is no benefit in MSFE terms from building models of disaggregates when only the aggregate is of interest. That analysis assumes that all conditioning information is available at $T-1$, and does not make use of the contemporaneous information available at T that is an essential component of the nowcasting strategy. Discarding information available at T is particularly costly when breaks occur simultaneously across the disaggregates (or a subset) as this information can be used to ensure robust nowcasts despite the breaks. This implies that either [II] or [IIIb] is a preferable strategy, especially given a changing set in $\mathbf{y}^0_{J_T,T}$.

As discussed in section 3.2, there is a range of methods for unbalanced panels in which all available data are used, including dynamic factor models, see Giannone et al. (2008) and Schumacher and Breitung (2008), and mixed frequency time-series models such as MIDAS, see Kuzin et al. (2009). Marcellino and Schumacher (2010) combine factor models with MIDAS, and Ferrara et al. (2010) use non-parametric methods, based on nearest neighbours and on radial basis function approaches to nowcast unbalanced monthly data sets. These methods do not take account of structural breaks in the disaggregates, instead abstracting from wide-sense non-stationarity. Hence, the approach we propose is to use [II], but augmented by additional contemporaneous variables and location shift detection, the issues to which we now turn.

3.6.2 Common features

The disaggregate series are likely to have common components, such as cycles and trends, as well as idiosyncratic elements, so it can be useful to include these in the GUM: see Espasa and Mayo-Burgos (2012). As N_T is large, it may be difficult to identify (say) cointegrating relationships across all disaggregates, but factors can proxy missing cointegration information. There are many ways to obtain the factors f_i for $i = 1, \ldots, q$: Stock and Watson (2002b) suggest using static principal component analysis applied to $\mathbf{y}_{d,t} = (\mathbf{y}_{1t}, \ldots, \mathbf{y}_{N_t})'$:

$$\mathbf{f}_t = \widehat{\mathbf{G}}' \mathbf{y}_{d,t}$$

where the set of q factor loadings is collected in $\widehat{\mathbf{G}}$, the $N_T \times q$ matrix of eigenvectors corresponding to the q largest eigenvalues of the sample covariance matrix $\widehat{\mathbf{\Omega}}_{y_d}$. Alternatively, Forni et al. (2005) propose a weighted version of the static principal components estimator, where time series are weighted according to their signal-to-noise ratio. Although there is no asymptotic theory for static principal components (PCs) computed from I(1) data, viewed as data-based orthogonal transforms, PCs seem to capture the main aspects of common trends and cycles, a result first described for aggregates by Stone (1947), so are probably sufficient for the purpose of nowcasting currently missing disaggregates. As the PCs mainly capture long-run relationships, the available information on the J_T known disaggregates at T will not contribute much, so the PCs can be calculated over the full sample to $T-1$.

In practice, we propose separating the disaggregates into blocks corresponding to groups of variables that exhibit common trends or cycles, as in section 3.7. The factors would be obtained for each block, f_i^j, where $i = 1, \ldots, q$ denotes the factor and $j = 1, \ldots, b$ denotes the block $\mathbf{y}_{d,j}$ where $\mathbf{y}_d$ is divided into b subsets, but all factors could be included in the GUM. A key innovation is that both individual explanatory variables and the factors are included jointly, in contrast to much of the dynamic factor models literature. This is feasible due to the ability of *Autometrics* to handle more variables than observations, and hence perfect collinearity, as described in section 3.5. If factors are helpful in explaining movements in the disaggregates, they should be retained in the selected model.

Changing collinearity in nowcasting

Following the analyses in Clements and Hendry (1998) and Clements and Hendry (2005), consider a linear, constant-parameter DGP where the first equation can be written for nowcasting as the contemporaneous model:

$$y_t = \mathbf{x}_t'\beta + \epsilon_t \tag{3.62}$$

where $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$, assuming the simplest marginal process for illustration:

$$\mathbf{x}_t \sim \mathsf{IN}_n\left[\mathbf{0}, \mathbf{\Sigma}\right] \tag{3.63}$$

independently of $\{\epsilon_t\}$, where $\mathbf{\Sigma} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}'$ with $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$ in-sample, and $\mathbf{z}_t = \mathbf{H}'\mathbf{x}_t$ so that:

$$\mathsf{E}\left[\mathbf{f}_t\mathbf{f}_t'\right] = \mathbf{H}'\mathsf{E}\left[\mathbf{x}_t\mathbf{x}_t'\right]\mathbf{H} = \mathbf{H}'\mathbf{\Sigma}\mathbf{H} = \mathbf{\Lambda}$$

leading to the principal components (or factor) representation of (3.62):

$$y_t = \mathbf{x}_t'\mathbf{H}\mathbf{H}'\beta + \epsilon_t = \mathbf{f}_t'\delta + \epsilon_t \tag{3.64}$$

For simplicity we ignore sampling variability in estimating the principal components.

After time T, $\mathbf{\Sigma}$ changes to $\mathbf{\Sigma}^*$:

$$\mathsf{E}\left[\mathbf{x}_{T+1}\mathbf{x}_{T+1}'\right] = \mathbf{\Sigma}^* = \mathbf{H}'\mathbf{\Lambda}^*\mathbf{H}$$

with a concomitant shift to:

$$\mathsf{E}\left[\mathbf{f}_{T+1}\mathbf{f}_{T+1}'\right] = \mathbf{\Lambda}^*.$$

The 1-step ahead MSFE, $\mathsf{E}[\widehat{\epsilon}_{T+1|T+1}^2]$, for known regressors, after estimating either (3.62) or (3.64) is:

$$\sigma_\epsilon^2\left(1 + T^{-1}\mathsf{E}\left[\mathbf{x}_{T+1}'\mathbf{\Sigma}^{-1}\mathbf{x}_{T+1}\right]\right) = \sigma_\epsilon^2\left(1 + T^{-1}\mathsf{E}\left[\mathbf{f}_{T+1}'\mathbf{\Lambda}^{-1}\mathbf{f}_{T+1}\right]\right) = \sigma_\epsilon^2\left(1 + T^{-1}\sum_{i=1}^{n}\frac{\lambda_i^*}{\lambda_i}\right) \tag{3.65}$$

In (3.63), let $\mathbf{x}_t' = (\mathbf{x}_{1,t}' : \mathbf{x}_{2,t}')$ of dimensions n_1 and $n_2 = n - n_1$ setting $\mathsf{E}\left[\mathbf{x}_{1,t}\mathbf{x}_{2,t}'\right] = \mathbf{\Sigma}_{12} = \mathbf{0}$ both to simplify the algebra and emphasize that the problem is not 'omitted variables bias'. Then:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \text{ and } \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{22} \end{pmatrix} \tag{3.66}$$

so that

$$\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11}\mathbf{\Lambda}_{11}\mathbf{H}_{11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{22}\mathbf{\Lambda}_{22}\mathbf{H}_{22}' \end{pmatrix}$$

which similarly must apply after the shift:

$$\begin{pmatrix} \mathbf{\Sigma}_{11}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22}^* \end{pmatrix} = \begin{pmatrix} \mathbf{H}_{11}\mathbf{\Lambda}_{11}^*\mathbf{H}_{11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{22}\mathbf{\Lambda}_{22}^*\mathbf{H}_{22}' \end{pmatrix}$$

Consider a nowcasting model that only includes the subset $\mathbf{x}_{1,t}$, leading to the nowcast:

$$\widetilde{y}_{T+1|T+1} = \mathbf{x}'_{1,T+1}\widetilde{\gamma}_1, \tag{3.67}$$

where:

$$\mathsf{E}\left[\widetilde{\gamma}_1\right] = \beta_1 = \gamma_1 \tag{3.68}$$

and:

$$\mathsf{V}\left[\widetilde{\gamma}_1\right] \approx T^{-1}\left(\sigma_\epsilon^2 + \beta_2'\mathbf{\Sigma}_{22}\beta_2\right)\mathbf{\Sigma}_{11}^{-1}.$$

The nowcast error $\widetilde{\epsilon}_{T+1|T+1} = y_{T+1} - \widetilde{y}_{T+1|T+1}$ is:

$$\widetilde{\epsilon}_{T+1|T+1} = \mathbf{x}'_{1,T+1}\left(\gamma_1 - \widetilde{\gamma}_1\right) + \mathbf{x}'_{2,T+1}\beta_2 + \epsilon_{T+1} \tag{3.69}$$

with unconditional expectation $\mathsf{E}\left[\widetilde{\epsilon}_{T+1|T+1}\right] = \mathbf{0}$ and approximate MSFE:

$$\begin{aligned}\mathsf{E}\left[\widetilde{\epsilon}^2_{T+1|T+1}\right] &\approx \sigma_\epsilon^2 + \mathsf{E}\left[\mathbf{x}'_{1,T+1}\mathsf{V}\left[\widetilde{\gamma}_1\right]\mathbf{x}_{1,T+1}\right] + \beta_2'\mathsf{E}\left[\mathbf{x}_{2,T+1}\mathbf{x}'_{2,T+1}\right]\beta_2 \\ &\approx \mathsf{E}\left[\widehat{\epsilon}^2_{T+1|T+1}\right] + T^{-1}\sigma_\epsilon^2\sum_{j=n_1+1}^{n}\left(\tau^2_{\gamma_j} - 1\right)\frac{\lambda_j^*}{\lambda_j}\end{aligned} \tag{3.70}$$

dropping the term of $\mathsf{O}(T^{-2})$, and using (3.65) where $\mathbf{H}_{22}\beta_2 = \gamma_2$ with:

$$\tau^2_{\gamma_j} = \frac{T\gamma_j^2\lambda_j}{\sigma_\epsilon^2},$$

which is the non-centrality of the t^2 test of the null that $\gamma_j = 0$ in (3.62) for $j = n_1 + 1, \ldots, n$.

Since only $\mathbf{x}_{1,t}$ is included in (3.67), when $\mathbf{\Sigma}_{12} = \mathbf{0}$ the corresponding factors must be $\mathbf{f}_{1,t} = \mathbf{H}_{11}\mathbf{x}_{1,t}$ so the orthogonalized estimated nowcasting model is:

$$\overline{y}_{T+1|T+1} = \mathbf{f}'_{1,T+1}\overline{\theta}_1$$

where $\mathbf{f}'_{2,T+1}\theta_2$ is omitted with:

$$\mathsf{V}\left[\overline{\theta}_1\right] \approx T^{-1}\left(\sigma_\epsilon^2 + \theta_2'\mathbf{\Lambda}_{22}\theta_2\right)\mathbf{\Lambda}_{11}^{-1}$$

and a nowcast error $\overline{\epsilon}_{T+1|T+1} = y_{T+1} - \overline{y}_{T+1|+1T}$:

$$\overline{\epsilon}_{T+1|T+1} = \mathbf{f}'_{1,T+1}\left(\delta_1 - \overline{\theta}_1\right) + \mathbf{f}'_{2,T+1}\theta_2 + \epsilon_{T+1} \tag{3.71}$$

having an unconditional expectation $\mathsf{E}\left[\overline{\epsilon}_{T+1|T+1}\right] = \mathbf{0}$ and approximate MSFE:

$$\begin{aligned}\mathsf{E}\left[\overline{\epsilon}^2_{T+1|T+1}\right] &\approx \sigma_\epsilon^2 + \mathsf{E}\left[\mathbf{f}'_{1,T+1}\mathsf{V}\left[\overline{\theta}_1\right]\mathbf{f}_{1,T+1}\right] + \theta_2'\mathsf{E}\left[\mathbf{f}_{2,T+1}\mathbf{f}'_{2,T+1}\right]\theta_2 \\ &\approx \mathsf{E}\left[\widehat{\epsilon}^2_{T+1|T+1}\right] + T^{-1}\sigma_\epsilon^2\sum_{j=n_1+1}^{n}\left(\tau^2_{\theta_j} - 1\right)\frac{\lambda_j^*}{\lambda_j}\end{aligned} \tag{3.72}$$

when:

$$\tau^2_{\theta_j} = \frac{T\theta_j^2\lambda_j}{\sigma_\epsilon^2}.$$

These non-centralities are orthogonal combinations of those in (3.70), and so could differ substantially. Nevertheless, the most important implications of changes in collinearity in forecasting are that an estimated model can have a smaller MSFE than the estimated DGP when omitted regressors (factors) have $\tau^2 \leq 1$ despite being mis-specified, and conversely all regressors with $\tau^2 > 1$ should be included in the forecasting equation. Changes in collinearity can markedly increase the MSFE when $\lambda_j^* >> \lambda_j$, irrespective of including or

excluding regressors with $\tau^2 \leq 1$, so omitting relevant collinear regressors or orthogonal factors need not improve a forecasting model. Finally, orthogonalizing transforms 'to reduce collinearity' are not useful when collinearity changes, since (3.70) and (3.72) both depend on the same eigenvalue ratios, albeit with different non-centrality deviations from unity.

3.6.3 Using all available information

The problem with nowcasts from lagged variables only is that additional information may be available through both other contemporaneous data and disaggregates known at T, $y_{i,T}$ for $i = 1, \ldots, J_T$, allowing for more rapid identification of outliers or location shifts. Hence, we propose an augmented nowcast denoted, $\widetilde{y}_{i,T|T-\delta}$.

There are many possible sources of contemporaneous data, including up-to-date values for some of the relevant disaggregates, rapidly and frequently observed outcomes for variables such as retail sales; correlated variables like road traffic and air passenger numbers or energy consumption; surveys of consumers and businesses about their plans and expectations; and more recent innovations including prediction markets and Google Trends: see Choi and Varian (2012). Models that exploit related series, possibly in combination with univariate time-series models, can help improve the accuracy with which any missing data are estimated, by adding the proxy as an explanatory variable in a model for the variable to be nowcast. The advantage of *Autometrics* is that all covariate information can be included in the general model at the outset, and the data characteristics will determine whether the explanatory variables are relevant historically in explaining the individual disaggregate series. The model selection for the individual nowcast models could be undertaken for every new release of data, which is feasible for automatic model selection. Hence, if a break is detected the model will be updated as soon as information on the break is available. The robustness of the model specification and parameter constancy can be tested alongside the break detection tests, but the distinction between internal and external breaks discussed in Castle, Fawcett, and Hendry (2010) points to different approaches in even those two cases.

Surveys

Survey information could be used directly to modify estimates of the nowcast origin values, or as possible additional regressors, or as part of a signal extraction approach to estimating missing data on the disaggregates, or as one of the devices to be pooled. We doubt their likely efficacy as leading indicators following the critiques in Diebold and Rudebusch (1991) and Emerson and Hendry (1996), since their *ex post* performance is usually superior to the *ex ante*, reflected in regular revisions of the indicator components of indexes.[8] However, we test whether surveys are able to provide timely detection of structural breaks in the empirical example.

Covariate information

Variables like retail sales are observed more frequently and rapidly than aggregates like GDP, and while they include some of the information needed for expenditure measures of GDP, they are also potentially correlated with other variables that are only available with greater latency. For example, retail sales are published approximately 32 days after the end of the reference period, whereas consumption data, for which retail sales is a reasonable proxy, is only released approximately 64 days after the end of the reference period along with the first full GDP release (excluding the flash estimate). Other proxies, such as new passenger car registrations, construction output and industrial new orders are also available more rapidly than components of the expenditure or income based measures of GDP.

Google query data

Choi and Varian (2012) show that *Google Trends* data can help improve their nowcasts of the current level of activity for a number of different US economic time series, including automobile, home, and retail sales, as well

[8] Note from the editor: We will adopt the term "indices" rather than "indexes" where applicable going forward.

as travel behaviour. They add the associated Google query variables to simple linear seasonal-AR models to measure the additional 'predictive' power, and although the resulting models are used for forecasting, their focus is on 'predicting the present'.

In an *Autometrics* approach, subsets could be used when nowcasting each missing disaggregates, or a small subset of what are deemed to be potentially relevant additional data series could be added to the information set for the aggregate under analysis: Doornik (2009b) illustrates this approach. Furthermore, the volumes of Google data available could enable measures of variance via the realized volatility of intra-day activity, and even complete distributions via a non-parametric approach. The optimal level of aggregation for such data is an empirical question, and mixed frequencies of data could draw on a MIDAS-type of approach: see, *inter alia*, Ghysels et al. (2007) and Clements and Galvão (2008). The selected model specifications could be maintained and updated within quarters, or the models could be re-selected intermittently if breaks have occurred.

Prediction market data

Prediction markets for the relevant measures could be another valuable source of ancillary information: see Wolfers and Zitzewitz (2004), Gil and Levitt (2007) and Croxson and Reade (2014) *inter alia*. These information markets, like Iowa Electronic Markets for political outcomes and Betfair for sporting competitions, are claimed to have more accurate predictions than polls, surveys, or expert judgement, and to have predicted high profile events well, such as the probability of President Obama's first victory, as in figure 3.18. However, it is unclear how well such markets forecast relative to robust econometric equations, as location shifts occur intermittently, even if the reported probabilities adjust rapidly to new information: see e.g., Hendry and Reade (2008a).

Figure 3.18: Prediction market based probabilities of an Obama victory.

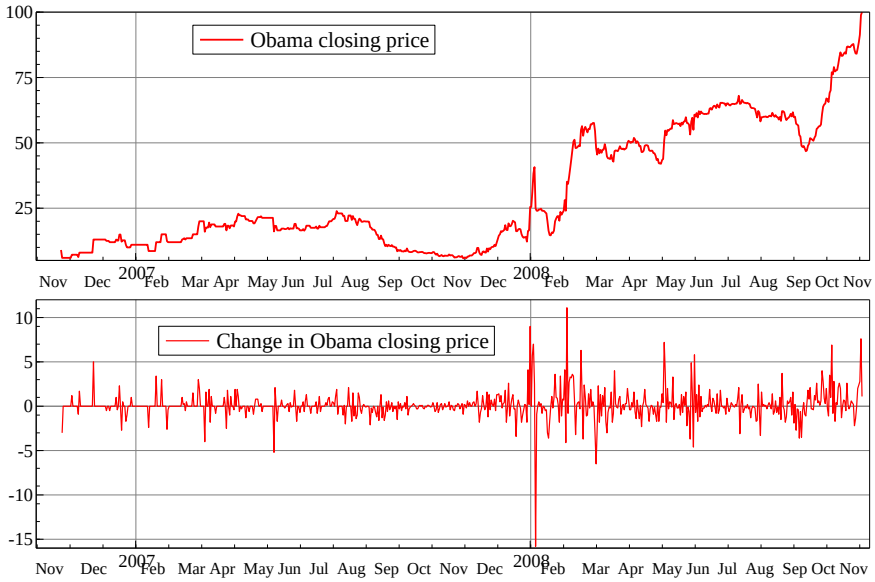


Figure 3.18 shows that early forecasts are very poor, leading to forecast failure at long horizons. Certainly punters acquire information from many sources, including the economy, but also from poll outcomes and opinion surveys. There are large jumps at major events, such as winning the nomination etc., but adjustment is not monotonic even after new shocks occur. Rapid updating seems to occur as news arrives, and accuracy improves at much shorter horizons, but even late in time, there are very inaccurate forecasts of the final probability. For example, the series predicts Obama's non-election as late as mid-September, some time after the structural break from Sarah Palin entering the race. Conversely, from mid-February onwards, most probabilities are above 0.5, so would be accurate for the outcome of an election victory if $p > 0.5$ was the decision

rule. Prediction markets inevitably have a termination point, by which date the probability has converged to either 0 or 1. Most graphs of the probabilities drift towards the actual outcome, but there can be 'last second' major breaks (e.g., an overtime goal in a soccer match), leading to dramatic switch in probabilities. Like most economic forecasts, future location shifts seem to be assumed absent.

Overall, however, as participants are risking their own money with their views, near their termination date, relevant prediction market real-time measures could be a useful addition to models measuring and nowcasting the present state of the economy. For example, the July 2009 Iowa Electronic Market was for bets on the Federal Reserve's Monetary Policy decisions, possibly revealing participants' implicit knowledge about the state of the US economy.

3.6.4 Seasonal adjustment

Seasonal adjustment (SA) of current (and recent past) data itself implicitly entails forecasting, as most SA procedures are long 2-sided moving-average filters, which have to be 'folded' at the end points of the sample, corresponding to extrapolating future data. Such 'forecasts' are often made by techniques that are different from the methods used for interpolating missing data in the disaggregated preliminary series. Conversely, if seasonally-adjusted disaggregated series underlie the published data, then the missing series must have been forecast in SA form. It is hard to ascertain whether such differences in methodology are of consequence relative to the other difficulties confronting the construction of preliminary data. Certainly, inconsistencies must arise when the forecasts implicit in the SA are not based on the same method as that used to fill in the missing data. Moreover, different outcomes will result if the data are first seasonally adjusted then forecast, or forecast then seasonally adjusted. Nevertheless, such errors are probably small relative to the mistakes in the initial data and any forecast numbers used to fill gaps.

3.6.5 Modeling the disaggregates

Models are required of all N_T disaggregates. As breaks may have occurred in the individual series over the sample period, it is important to account for location shifts in these models and we do so using IIS as described in section 3.5.3. An illustrative in-sample general unrestricted model (GUM) for modeling the disaggregates is:

$$y_{i,t} = \gamma' \mathbf{x}_{t-\delta} + \theta' \mathbf{f}_{t-1} + \sum_{j=1}^{T-1} \varsigma_j 1_{t_j=t} + \rho_i y_{i,t-1} + \nu_t, \tag{3.73}$$

for $t = 1, \ldots, T-1$, and $i = 1, \ldots, N_T$. $\mathbf{x}_{t-\delta}$ denotes all the available information including other relevant lagged disaggregates and additional explanatory variables such as survey information, leading indicators etc. (as discussed in section 3.6.3), $\mathbf{f}_{t-1}$ denotes a set of q latent common factors (discussed in section 3.6.2), and $1_{t_j=t}$ is the set of $T-1$ individual impulse indicators taking the value unity at t and zero otherwise. Selection is undertaken using *Autometrics* on each series' available sample (possibly $t = 1, \ldots, T-1$, but potentially shorter for recently introduced entities), and the resulting estimates are:

$$\widehat{y}_{i,t} = \widehat{\gamma}' \mathbf{x}_t^* + \widehat{\theta}' \mathbf{f}_{t-1}^* + \widehat{\rho}_i y_{i,t-1} + \widehat{\varsigma}' \mathbf{d}_t \tag{3.74}$$

where $\mathbf{x}_t^*$ and $\mathbf{f}_{t-1}^*$ denote the retained variables and factors after selection and $\mathbf{d}_t$ collects the retained impulse indicators from selection of (3.73). Loose significance levels could be used to select the forecasting model regressors in $\mathbf{x}_t^*$: Clements and Hendry (2005) show that all regressors with squared non-centralities of their t-statistics greater than unity should be retained, although it has proved difficult to link that result to an empirical selection rule (see section 3.6.2). However, tight significance levels should be used to select the impulse indicators as $\alpha(T-1)$ will be retained on average under the null from (3.73). In practice, *Autometrics* only imposes one significance level, so selection could be undertaken using a 2-step procedure by appealing to the Frisch and Waugh (1933) theorem, first identifying the impulse indicators and then selecting the regressors conditional on them.

3.6.6 Nowcasting strategy

We can now integrate all the components of the Nowcasting strategy. Standing at time T, with all the available data for $T-1$ and any selected higher-frequency recent and contemporaneous data, the models in (3.73) have been specified, selected and estimated as in (3.74). From these, forecasts of the outcome at T can be developed, either directly if there were no recent known location shifts, or robustly otherwise. We first take the former:

$$\widetilde{y}_{i,T|T-\delta} = \widehat{\gamma}'\mathbf{x}^*_{T-\delta} + \widehat{\theta}'\mathbf{f}^*_{T-\delta} + \widehat{\rho}_i y_{i,T-1} + \widehat{\varsigma}' d_{T-1} \tag{3.75}$$

for $i = 1, \ldots, N_T$, where $\delta \simeq 0$. A 'flash forecast' of the aggregate y_T could be produced, and compared to other forecasts for early warning signs. That is why we suggest also calculating the robust forecast based on the first difference of (3.75) but without the differenced shift indicators:

$$\Delta\overline{y}_{i,T|T-\delta} = \widehat{\gamma}'\Delta\mathbf{x}^*_{T-\delta} + \widehat{\theta}'\Delta\mathbf{f}^*_{T-\delta} + \widehat{\rho}_i \Delta y_{i,T-1} \tag{3.76}$$

where the parameter estimates are unchanged, and $\overline{y}_{i,T|T-\delta} = \Delta\overline{y}_{i,T|T-\delta} + y_{i,T-1}$. However, robust methods do not forecast location shifts, merely adapt rapidly after their occurrence, so have the same large forecast error when a break happens after a forecast is made. Thus, it is only necessary to switch after a break has been observed. A decision on that occurrence can be based on testing whether $\widehat{\varsigma}_i$ was significant, using a relatively conservative significance level like 1.0%, noting that conventional tests have low or no power at sample end-points.

Once the first J_T components are observed, denoted $\widehat{y}_{i,T|T}$, their forecasts can be compared with these outcomes to evaluate forecast accuracy. Denote the forecast errors by:

$$\widehat{e}_{i,T|T} = \widehat{y}_{i,T|T} - \widetilde{y}_{i,T|T-\delta}, \qquad i = 1, \ldots, J_T,$$

Large values of $\widehat{e}_{i,T|T}$ relative to the standard error of the corresponding equation in (3.74) suggest recent location shifts, which then have implications for as yet unobserved cognate, or related, variables. Major discrepancies between the forecast and the measured diasaggregates suggest a need to rapidly update those for associated location shifts.

Because the J_T observed variables can change every period, they cannot be introduced as regressors at the model selection stage in (3.73), so we now discuss how $g(\cdot)$ in (3.54) is formed. Our approach is to use the information from the $\widehat{e}_{i,T|T}$ to modify the forecasts of the $N_T - J_T$ unknown disaggregates from (3.75), albeit there may remain a role for judgment, depending on whether the forecast errors for the measured J_T variables are thought to be measurement mistakes (and hence transitory) or location shifts, and whether idiosyncratic or common and so correlated with the unknown disaggregates at T. Partition the J_T observed variables into related sets (e.g., automobile related, industrial production, prices, financial variables, interest rates, labour variables, housing market variables, leisure services, etc.), where there are common trends or cycles within subsets of disaggregates and hence close linkages, so breaks are likely to spread within blocks. Denote these groups by $J^*_{k,T}$ for $k = 1, \ldots, K_T$ say. Within each subset, calculate the average $\widehat{e}_{i,T|T}$ (possibly as a proportion of the level of the series at $T-1$) and denote this by $\overline{e}^*_{k,T|T}$. Then we propose modifying the forecasting rule after substantial breaks for each partition by:

$$\widehat{y}_{h,T|T} = \widetilde{y}_{h,T|T-\delta} + \overline{e}^*_{h,T|T}, \quad h \in J^*_{k,T} \tag{3.77}$$

In essence, (3.77) is an intercept-corrected (IC) forecast analogous to setting the forecast 'back on track', but across disaggregate series as opposed to through time, which is the usual procedure. Consequently, in contrast to standard IC procedures that double the error variance, only a small variance increase should occur from (3.77) when each $J^*_{k,T}$ contains a number of series, yet potentially offsetting an important shift. This strategy is computationally feasible, allows for breaks in-sample and at the nowcast origin, while employing the full information set, but relies on the known disaggregates containing a correlated signal about the unknown disaggregates.

An alternative after the detection of a location shift is to robustify the forecasts of the missing series directly by using (3.76). However, while this will exploit the knowledge that a location shift has occurred, it will not use the information on its sign and magnitude provided by $\overline{e}^*_{k,T|T}$. That would not matter greatly for a shift that happened in an earlier period, but could be a substantive loss for a shift during $T-1$ to T. Such a robust device is akin to a 'Holt–Winters in-fill', but on the differences relative to the usual application, and while it does not impose the break magnitude as in (3.77), it doubles the innovation error variance. Both approaches could be used and pooled, there being considerable evidence of the benefits of pooling good methods, although not on 'pools' where some bad methods are included: see Hendry and Reade (2008b).

The final stage is to aggregate the nowcast series with data known at T as follows:

$$\widetilde{y}_{T|T} = \sum_{i=1}^{J_T} w_{i,T} \widehat{y}_{T|T} + \sum_{i=J_T+1}^{N_T} w_{i,T} \widetilde{y}_{i,T|T-\delta} \tag{3.78}$$

This approach of utilizing data on the subset of available disaggregates to construct the desired aggregate addresses the 'missing data', 'changing database', and 'break' problems, as impulse-indicator saturation removes past breaks, as well as including the additional contemporaneous variables. *Autometrics* writes the *Ox* code for each case analyzed, so it is feasible to develop a general program for application to many variables.

3.6.7 Some simulation evidence

Simulation experiments of the alternative nowcasting approaches are reported in Kitov (2012). Here we briefly summarize the results in relation to the performance of the robust and intercept-corrected devices for permanent and transitory location shifts. The parameters in the simulations are summarized in table 3.1 for a VAR(1) DGP in 2 disaggregates. *Autometrics* selected the relevant subset of variables from the 100 initially included while also implementing IIS.

Table 3.1: Parameters in the simulations.

$M = 1000$	number of simulated replications
$n = 100$	total number of exogenous candidate variables
$L = 10$	number of relevant exogenous explanatory variables
$T = 200$	pseudo in-sample sample size, forecast origin
$y_{i,0} = 0$	initial values for two disaggregates
$h = 4$	pseudo out-of-sample sample size
$5\sigma_{y_t}$	magnitude of the location shift
$\epsilon_{i,t} \sim \mathsf{IN}[0,1]$	DGP errors, for $i = 1, 2$

The results showed that intercept-corrected nowcasts can result in almost full adjustment for the bias created by a common location shift when contemporaneous information is used for correction. However, when the break is not transmitted forward, the intercept correction results in biased nowcasts that may differ by a large magnitude from those of the conditional model. Consequently, intercept correction transpires to be a high risk–high reward device, so should be used only when there is sufficient evidence of relatively permanent co-breaking in the series.

The robust nowcasts performed well on average, resulting in unbiased predictions under more scenarios than other methods. They led to a lower risk of getting the nowcasts badly wrong when the adjustment was unnecessary, yet the correction margins based on RMSE or mean absolute error (MAE) measures were much lower than those for the intercept-corrected nowcast in cases when both methods resulted in an improvement relative to the conditional model. Moreover, the robust device can successfully correct the nowcast in the worst-case scenario when breaks in series happen to be of opposite directions, as when one variable is substituted for another, explaining its sudden fall, for example.

3.6.8 Mis-measured disaggregates and location shifts

The most difficult setting is when the reported disaggregates are mis-measured at the same time as a location shift occurs. The former can happen because an unrepresentative sample of reports is returned to the statistical agency (e.g., large companies doing well when small are unknowingly doing badly or vice versa, etc.). If contemporaneously a location shift has occurred, inferring the missing disaggregates can be problematic. We address this problem by comparing the two forecasting devices with the measured disaggregate outcomes. A further useful comparator can be the change from the previous observation for each disaggregate. There are five possible outcomes.

a] All three measures deliver the same general outcome. The most probable state is that the data disaggregates are accurate and there is no break, but an undetectable contemporaneous offsetting of mis-measured data by a similar magnitude shift in the opposite direction is conceivable. The missing data can best be in-filled using the average of the two forecasting devices.

b] The recorded data and the regression forecast are similar, but the robust device differs from both. This reveals that a location shift or a serious mis-measurement had happened in the previous period, but either the former was not carried forward (i.e., probably an earlier mis-measurement), or a shift has indeed happened but been mis-measured this period (noting that the regression forecast may not reflect a previous location shift). If investigation cannot clarify which is the correct state before the release date, the best that can be done is probably to in-fill using the regression forecast.

c] The recorded disaggregate data and the robust device are similar, but the regression forecast differs from both. This probably reveals that a location shift occurred in the previous period and the data reflect that but the contemporaneous variables did not correct the regression forecast enough. The missing data can best be in-filled using the robust forecasting device.

d] The robust device and the regression forecast are similar, but both differ from the recorded disaggregate data. If a location shift had occurred in the previous period, then it must have been captured by the contemporaneous variables in the regression forecast in order for it to match the robust device, but the most likely scenario is that either a new shift has happened that is not reflected in either forecast, or there is a serious mis-measurement. If the last can be excluded, the missing data can best be in-filled using the average of the two forecasting devices corrected by the average forecast error from the recorded disaggregates.

e] All three deliver different outcomes. Clearly the state of nature must have changed from the previous period, and had also changed then. Adjusting the robust device by the average forecast error from the recorded disaggregates may be the best that can be achieved if the situation cannot be clarified in time.

The next section presents an illustrative application of the nowcasting strategy.

3.7 Nowcasting UK GDP growth

This final section presents an empirical application of the nowcasting methodology to UK GDP growth over the last two decades. The apparent break in the GDP growth series, represented by the recession, is a good candidate for testing the nowcasting performance of *Autometrics* with IIS for break correction, given the recent poor performance of the GDP estimates produced by the ONS. In addition, it is interesting to check whether *Autometrics* can provide robust nowcasts for two periods which had different levels of variability. To that end, the period spanning the recent turmoil is compared with a calmer period during the 1990s.

The conditioning set includes many monthly leading indicators, as well as preliminary GDP growth vintages. The GDP growth nowcasts are computed using bridge equations. To check for potential benefits of information about common breaks, leading indicators are grouped in blocks of correlated variables that are more likely to break simultaneously. *Autometrics* model selection with IIS then produces forecasts of the leading indicators for the missing observations at the end of the sample. A robust forecasting device for monthly variables is also

tested. An analysis of common breaks within blocks of closely linked variables checks whether contemporaneous intercept correction could be applied to improve nowcasting performance. Finally, nowcasting models for GDP growth are selected using *Autometrics*.

Section 3.7.1 reviews the data, section 3.7.2 describes the handling of the leading indicators, section 3.7.3 outlines the nowcasting methodology, and section 3.7.4 discusses the model specifications. Then sub-sections 3.7.5–3.7.7 discuss the estimates for vintage models, single-indicator models and augmented models respectively. Section 3.7.8 investigates contemporaneous breaks.

3.7.1 Data

The GDP growth series was downloaded from the ONS website,[9] and contains 76 quarterly observations for the period between 1993Q1 and 2009Q4. Seasonally-adjusted chained volume measures of GDP at constant 2008 prices were used. The latest GDP growth estimates, presumed to be the most accurate reported value to date, are the target variable, denoted by Δy_t, where y_t is the log of the GDP level and Δ is the first difference of log(GDP). The latest GDP growth vintage is an appropriate measure when the objective is to nowcast the actual change in output, as in Corradi et al. (2009) and Clements and Galvão (2009).

Four additional estimates of GDP growth were obtained, all from the GDP Revisions Triangle Release (ABMI) reported by the ONS. These are three preliminary vintages: $\Delta y_{t_q}^{v_1}$, $\Delta y_{t_q}^{v_2}$, $\Delta y_{t_q}^{v_3}$ released in the months directly following the reference quarter, and $\Delta y_{t_q}^{v_f}$ corresponding to the 'final estimate' reported three years after the end of the reference quarter (further revisions in later periods produce the latest estimate Δy_{t_q}). Consequently, the last is the most accurate measure of GDP available to date. A summary of the GDP estimates, their release dates and information content is provided in table 3.1.

Table 3.1: GDP growth estimates, release dates and data completeness.

Variable	Estimate	Release lag	Real data
$\Delta y_{t_q}^{v_1}$	Preliminary, first estimate	3.5 weeks	44%
$\Delta y_{t_q}^{v_2}$	Output, Income and Expenditure, second estimate	8 weeks	67%
$\Delta y_{t_q}^{v_3}$	UK National Accounts, first final estimate	12 weeks	80%
$\Delta y_{t_q}^{v_f}$	Final estimate 3 years later	3 years	
Δy_{t_q}	Latest available, most accurate estimate to date		

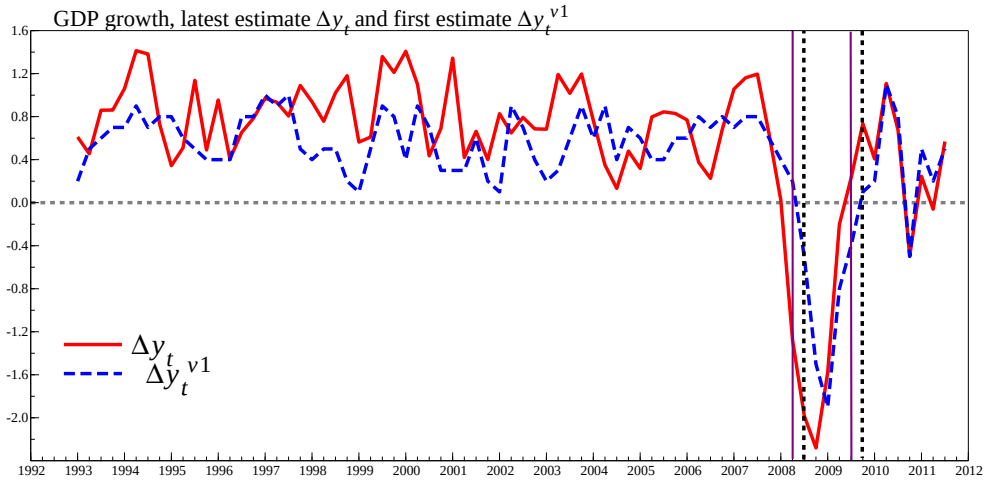
To see the differences between the preliminary and the final growth estimates, the least complete estimate, $\Delta y_{t_q}^{v_1}$, and the latest available one, Δy_{t_q}, are plotted in Figure 3.19. Although the patterns of growth are similar on average, the deviations between the two series are considerable, especially in periods of major changes. Until the start of the recession in 2008, GDP growth had been positive and relatively stable, with non-trending and mean-reverting properties. The two series nearly coincide after 2010 as preliminary measures have not yet been substantially revised, so the latest quarterly estimate deemed reliable enough here is 2009Q4.

The first instance of negative growth was reported in 2008Q2, according to Δy_{t_q}. However, the recession was missed by preliminary estimates at all real-time vintages by one quarter. According to these measures, including $\Delta y_{t_q}^{v_1}$, the recession did not start until 2008Q3, as can been seen on the plot. Thus, measurement errors during the recession were considerably larger than before. The recession started earlier than ONS reported for three consecutive National Accounts releases, and was only identified *ex post*. The measurement procedures failed to produce correct estimates of the output decline, with an estimation error of -1.5% at the first reported quarter of negative growth.

Further analysis of the revision process is presented in Figure 3.20, which depicts the differences between Δy_{t_q} and the three preliminary vintages. The three revision processes are close, even for the final estimate $\Delta y_{t_q}^{v_f}$, noting that $\Delta y_{t_q}^{v_f}$ is only available until 2008Q3 and for 2008Q1–2008Q2 is equal to the latest estimate.

[9] http://www.ons.gov.uk/ons/datasets-and-tables/index.html: this was done during the first half of 2012.

Figure 3.19: GDP growth in percentage terms based on the latest estimate Δy_t and the first preliminary estimate $\Delta y_t^{v_1}$ between 1993 and 2011.



All primary estimates tend to underestimate large positive levels of Δy_{t_q}, and to overestimate large negative ones, which happens in 70% of the cases. In general, even $\Delta y_{t_q}^{v_f}$ tends to underestimate the latest available estimate up to 2008Q3. The largest deviations from the most accurate estimates are reported during the recession, with two instances of 1.5% errors for the preliminary estimates.

Table 3.2: Revision to the latest and initial GDP growth estimate, mean error (ME) and standard deviation (SD) in percentage points for 1993Q1–2008Q3.

	$\Delta y_t^{v_1} - \Delta y_t$	$\Delta y_t^{v_3} - \Delta y_t$	$\Delta y_t^{v_f} - \Delta y_t$
ME (%)	-0.152	-0.149	-0.113
SD (%)	0.474	0.435	0.309

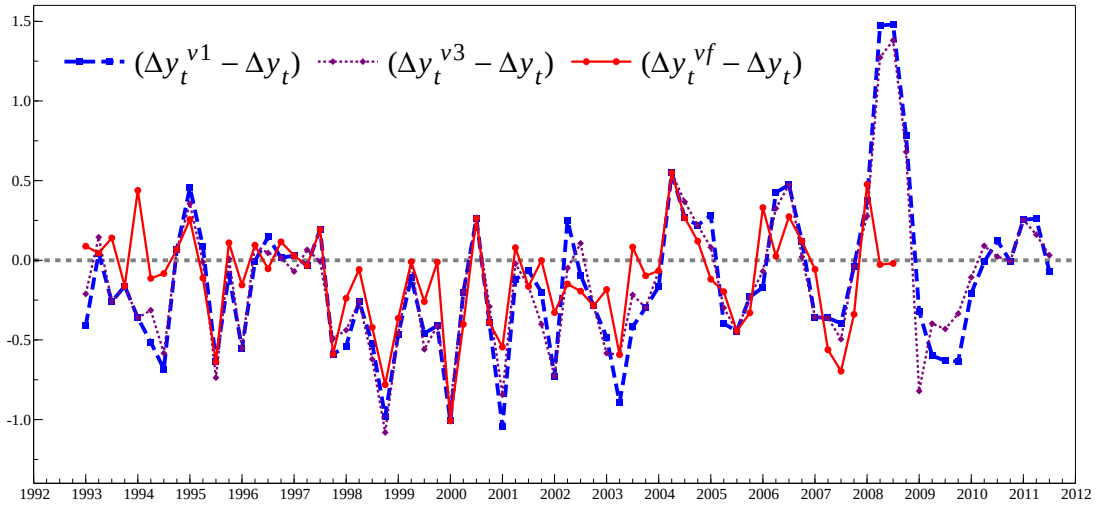
	$\Delta y_t^{v_2} - \Delta y_t^{v_1}$	$\Delta y_t^{v_3} - \Delta y_t^{v_1}$	$\Delta y_t^{v_f} - \Delta y_t^{v_1}$
ME (%)	-0.011	0.003	0.040
SD (%)	0.083	0.116	0.379

Finally, table 3.2 reports average errors and standard deviations for the revisions compared to Δy_t, revealing systematic underestimation on average for all vintages, even including the most recent overestimation during the recession, which drives the overall mean error up from -0.2. As shown in Figure 3.20, real-time vintages are not revised greatly compared to $\Delta y_t^{v_1}$. A noticeable, but unsystematic error, can be also seen for the final estimate, especially at the first instance of recession.

3.7.2 Leading indicators

The data set on the monthly disaggregated variables contains 60 series, over 1980Q1–2010Q2, and was kindly provided by Christian Schumacher: see Kuzin et al. (2012). All of these series had some merit in explaining GDP in previous studies, and include UK survey data, industrial production, labor market and employment statistics, trade variables, financial indicators and the corresponding series for the US and Europe. Leading indicators are released with different monthly lags relative to the current month t_m, where the actual days of publication differ throughout the month. In a real-time exercise, the information set can expand by a single variable and each new release can be treated as a nowcast origin, as in Bánbura et al. (2011). However, it suffices to fix the release lags relative to the current period for each variable, resulting in a constant structure

Figure 3.20: Revision process to the latest estimate of GDP growth, Δy_t, as compared with the $\Delta y_t^{v_1}$, $\Delta y_t^{v_3}$ and a revised GDP measure three years later, $\Delta y_t^{v_f}$.



for an unbalanced panel, where the pattern of missing observations at each point in time is the same. This produces a pseudo real-time data set, which in real time is only available at the end of each month.

The level series are non-stationary so data transformations are applied to produce variables that appear to be approximately I(0). For application of the common-break-correction devices, the leading indicators were grouped into 6 blocks, as in Castle, Fawcett, and Hendry (2011), where the blocks should contain variables with varying release lag orders to aid timely detection of shifts. Such variables should be highly correlated, increasing the likelihood of common breaks. Therefore, six blocks were formed by optimizing cross-correlations between the transformed series, so that within-block correlations are maximized, whereas between-blocks correlations are minimized, while trying to group variables with different release lags. The latter restriction is not always satisfied. For example, block 5 only contains financial variables that are available contemporaneously, but have negligible correlations with other variables and are therefore grouped separately.

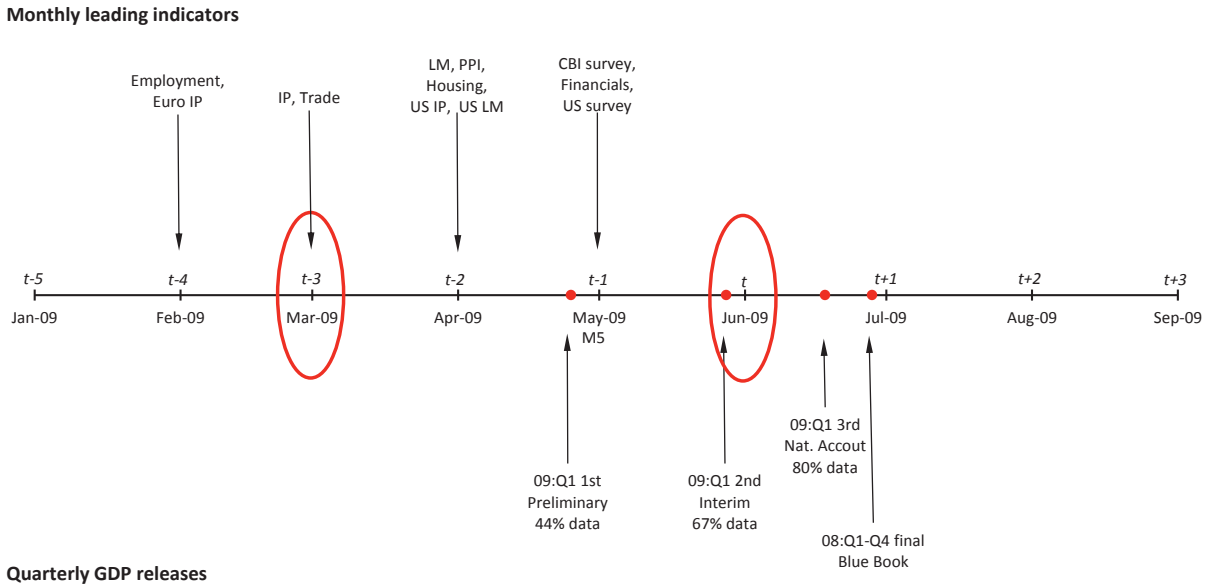
The actual timings of data releases for the leading indicators and GDP estimates are presented in Figure 3.21. Suppose that the current period $t = t_{m_6} = t_{q_3}$ is June. Then, as can be seen in the upper half of the figure, the latest UK and US soft data together with the financial variables are available for the previous month, i.e., with a one-month lag. The remaining leading indicators are released with longer lags. For instance, employment data and European industrial production are only available for February. Three real-time vintages of GDP growth are also shown in the figure, and for the illustrated example are all available for the first quarter.

3.7.3 Nowcasting methodology

Nowcasting GDP growth is performed using the bridge equations framework. At the first stage, direct forecasts of the missing observations for all leading indicators are computed within blocks of correlated variables. The forecasts are produced for the last 24 months of each corresponding sub-sample, that is for 1999M1–2000M12 and 2008M1–2009M12 respectively. Individual models are selected recursively using *Autometrics* with IIS at a 1% significance level. For a monthly variable z_{i,t_m} in block $k = \{1, \ldots, 6\}$, available with a lag l_i, such that the latest observation is z_{i,t_m-l_i}, the following GUM fully saturated by impulse indicators is formed from variables of the same block available in-sample:

$$z_{i,t_m} = \sum_{i=1}^{b_k} \sum_{j=1}^{12} \beta_{i,j} z_{i,t_m-l_i-j} + \sum_{t=1}^{t_m} \zeta_{i,t} 1_{i,t} \tag{3.79}$$

Figure 3.21: Pseudo release timing and structure of data vintages for the quarterly GDP measurements and monthly leading indicators.



where b_k is the total number of variables in block k and 12 is the longest lag entering the GUM. All monthly indicators are forecast within blocks, that is for $1 \leq i \leq b_k$. One, two or three periods ahead forecasts are computed depending on the corresponding release lag $l_i \in \{1, 2, 3\}$ to correct for the ragged edges. The earliest lag allowed to enter the GUM varies among variables and corresponds to the release delay l_i, so that all information that would have been available in real-time at the forecast origin is included, so contemporaneous observations for variables available at t_m are included in a model for all other series observed with a lag, but not vice versa. Forecasting models are also estimated for the variables released with no delays, resulting in one step ahead forecasts. Break detection is performed simultaneously with model selection and significant contemporaneous impulse indicators are retained. This is done for a later analysis of common breaks at the nowcast origin.

The second stage of the nowcasting procedure involves the formulation of the bridge equation linking the quarterly GDP growth with the monthly leading indicators. In order to check whether monthly inflow of information can monotonically improve nowcasting accuracy, three nowcast origins are compared. These are: $h^1_{t_{q_1}}$ corresponding to the second month of the quarter, $h^2_{t_{q_1}}$ corresponding to the third month of the quarter and $h^3_{t_{q_1}}$ one month after the reference quarter. The structure of nowcast vintages is shown in table 3.3, along with the releases of the real-time quarterly GDP growth estimates. Note that at $h^3_{t_{q_1}}$ the first preliminary estimate is available. The latter is intentionally included to test whether the flash estimates contain relevant information about actual growth in real time.

Provided that the period before the recession is characterized by mean reverting growth with low volatility, it is conjectured that the model selected for this sub-sample will differ substantially from the model selected for the recession. In order to check this, the whole available period is divided into two sub-samples. The first sub-sample contains 32 observations and spans the period between 1993Q1 and 2000Q4. The second sub-sample covers 36 observations over 2001Q1–2009Q4. For each period, the last eight quarters are nowcast as noted above.

The mixed-frequency data issue is tackled explicitly within the bridge equations framework by transforming the unmodeled monthly variables to produce a consistent quarterly panel so that a nowcasting regression for the quarterly data could be specified directly. Consequently, to use the full information content in the monthly variables, they are transformed into three distinct quarterly series, as outlined in Castle, Fawcett, and Hendry (2009). For every $\mathbf{z}_{t_m} = (z_{t_m}, \ldots, z_0)'$, denote the latest quarter t_q for which an observation $\mathbf{z}_{t_q}$ is available

Table 3.3: Nowcast horizons and GDP releases in real time. Each entry corresponds to a nowcast/release of the quarterly GDP growth in a quarter t_{q_i} available in month t_{m_j}.

	t_{m_1}	t_{m_2}	t_{m_3}	t_{m_4}	t_{m_5}	t_{m_6}	t_{m_7}	t_{m_8}	t_{m_9}	$t_{m_{10}}$	$t_{m_{11}}$	$t_{m_{12}}$
t_{q_1}		$h^1_{t_{q_1}}$	$h^2_{t_{q_1}}$	$h^3_{t_{q_1}}$ $y^{v_1}_{t_{q_1}}$	$y^{v_2}_{t_{q_1}}$	$y^{v_3}_{t_{q_1}}$						
t_{q_2}					$h^1_{t_{q_2}}$	$h^2_{t_{q_2}}$	$h^3_{t_{q_2}}$ $y^{v_1}_{t_{q_2}}$	$y^{v_2}_{t_{q_2}}$	$y^{v_3}_{t_{q_2}}$			
t_{q_3}								$h^1_{t_{q_3}}$	$h^2_{t_{q_3}}$	$h^3_{t_{q_3}}$ $y^{v_1}_{t_{q_3}}$	$y^{v_2}_{t_{q_3}}$	$y^{v_3}_{t_{q_3}}$
t_{q_4}	$h^3_{t_{q_4}}$ $y^{v_1}_{t_{q_4}}$	$y^{v_2}_{t_{q_4}}$	$y^{v_3}_{t_{q_4}}$								$h^1_{t_{q_4}}$	$h^2_{t_{q_4}}$

by τ, such that $\tau = t_q \leq t_m$, then:

$$\begin{aligned} \text{first month: } \mathbf{z}^1_{t_q} &= z_{\tau-2}, z_{\tau-5}, \dots \\ \text{second month: } \mathbf{z}^2_{t_q} &= z_{\tau-1}, z_{\tau-4}, \dots \\ \text{third month: } \mathbf{z}^3_{t_q} &= z_{\tau}, z_{\tau-3}, \dots \end{aligned} \tag{3.80}$$

where $\mathbf{z}^1_{t_q}$, $\mathbf{z}^2_{t_q}$ and $\mathbf{z}^3_{t_q}$ are vectors containing observations for the first, second and third months of the quarter respectively. This transformation results in a total of 180 quarterly leading indicators, in contrast with the approach of aggregating monthly data into quarterly indicators: see for instance Ferrara et al. (2010).

To see how the availability of the quarterly indicators changes with respect to the nowcasting horizon, table 3.4 reports the relative timing for a particular example of the Q3 nowcasts at three different horizons, where leading indicator z_{t_q} is available with a lag of 3 months.

Table 3.4: Earliest available quarterly indicators relative to nowcast horizons.

Month	t_{m_5}	t_{m_6}	t_{m_7}	t_{m_8}	t_{m_9}	$t_{m_{10}}$	$t_{m_{11}}$	$t_{m_{12}}$
Horizon				$h^1_{t_{q_3}}$	$h^2_{t_{q_3}}$	$h^3_{t_{q_3}}$		
$h^1_{t_{q_3}}$	$z^2_{t_{q_3}-1}$							
$h^2_{t_{q_3}}$	$z^2_{t_{q_3}-1}$	$z^3_{t_{q_3}-1}$						
$h^3_{t_{q_3}}$	$z^2_{t_{q_3}-1}$	$z^3_{t_{q_3}-1}$	$z^1_{t_{q_3}}$					

The farthest available observation to the right in each row shows the last observation available at each nowcast horizon and corresponds to the release month shown in the top row. For instance, at $h^1_{t_{q_3}}$, when the first nowcast for Q3 is produced at t_{m_8}, the indicator observations are available until time $t_{m_8} - 3 = t_{m_5}$. This corresponds to the quarterly series $z^2_{t_{q_3}-1}$, which is observed in the second month of Q2. For the last horizon, respectively, the latest available observation is $z^1_{t_{q_3}}$. A similar lower triangular release structure is applicable to variables reported with one and two month lags. Finally, the bridge equations for the quarterly GDP nowcasts are formed using the transformed quarterly series, which effectively solves the mixed frequency problem.

3.7.4 Model specifications

We now undertake a quasi real-time nowcasting exercise. All estimated models contain information that would have been available at the end of the month corresponding to each nowcast origin. For every horizon and every quarter, models are selected recursively using *Autometrics* with IIS at a 1% significance level, producing eight quarterly specifications for each subsample. Note that the nowcasting exercise is recursive in relation to the completeness of the information set. At each stage, the previous model is extended by additional data.

Therefore, each new GUM nests the previous one. Each model is constructed using the transformed quarterly series for the leading indicators, their lags and a full set of impulse indicators. This implies that the number of candidate variables exceeds the number of available observations. *Autometrics* can efficiently handle that issue by using a block search algorithm for model selection. The automatic model selection is then performed for the following specifications.

Vintage models (VM): a benchmark univariate model that includes the quarterly data on real time GDP growth vintages. In addition to the most accurate data at a real vintage v_3, this specification includes the earlier estimates, thereby utilizing the information contained in the revision process. It is noted that although the vintages are highly collinear, *Autometrics* is able to handle correlated data with satisfactory selection and forecasting performance, as was demonstrated by the simulation exercise above. To utilize all information available in real time, the GUMs differ for three nowcasting horizons and are shown in equation (3.81):

$$\begin{aligned} h^1_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q-1}, \ldots, \Delta y^{v_1}_{t_q-4}; \Delta y^{v_2}_{t_q-1}, \ldots, \Delta y^{v_2}_{t_q-4}; \Delta y^{v_3}_{t_q-2}, \ldots, \Delta y^{v_3}_{t_q-4}\right) \\ h^2_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q-1}, \ldots, \Delta y^{v_1}_{t_q-4}; \Delta y^{v_2}_{t_q-1}, \ldots, \Delta y^{v_2}_{t_q-4}; \Delta y^{v_3}_{t_q-1}, \ldots, \Delta y^{v_3}_{t_q-4}\right) \\ h^3_{t_q} : \Delta\hat{y}_{t_q} &= f\left(\Delta y^{v_1}_{t_q}, \ldots, \Delta y^{v_1}_{t_q-4}; \Delta y^{v_2}_{t_q-1}, \ldots, \Delta y^{v_2}_{t_q-4}; \Delta y^{v_3}_{t_q-1}, \ldots, \Delta y^{v_3}_{t_q-4}\right) \end{aligned} \tag{3.81}$$

At h^1_t , the most recent estimate of $\Delta y^{v_3}_{t_q-1}$ is still not available in real time and is therefore excluded from the GUM. The last horizon h^3_t overlaps with the release period for the flash estimate and thus the first contemporaneous vintage of GDP is included in that GUM.

Single-indicator in-sample models (SM): for preliminary assessment of each predictor's individual relevance, growth is projected onto the VM models augmented by one single monthly indicator z^j_{i,t_q}, resulting in a total of 60 models for each evaluation period, corresponding to every leading indicator in the information set. Since no block separation is necessary here, the forecasts for the missing observation are omitted, utilizing in-sample information only. The maximum lag for the indicators is 12 and the minimum lag corresponds the earliest observed realization. Therefore the GUMs for single models vary depending on the nowcasting horizon and the corresponding availability of the observations. The relevant lags are selected by *Autometrics*. Since these models should merely provide an indication of individual predictors' robustness and relevance for nowcasting growth, the actual results for these models will not be presented, and instead assessed relative to the VM model.

Augmented model with in-sample indicators (AI): the growth vintages from (3.81) are augmented with the full set of leading indicators. Only in-sample information enters the GUM so that the ragged-edge problem is not corrected for. The GUM for the AI model is effectively a combination of the VM and SM specifications. The nowcasting model from the selected specification then has the following form:

$$\Delta\hat{y}^{h^k_q}_{t_q} = \mathbf{m}\hat{\alpha}'\Delta\hat{\mathbf{y}}^{v}_{t_q} + \mathbf{m}\hat{\beta}'\hat{\mathbf{z}}_{t_q} + \mathbf{m}\hat{\zeta}'\hat{\mathbf{d}}_{t_q} \tag{3.82}$$

where $\Delta\hat{\mathbf{y}}^{v}_{t_q}$ is a vector containing the selected lags of the growth estimates, $\hat{\mathbf{z}}_{t_q}$ is a vector of lags of the relevant in-sample leading indicators and $\hat{\mathbf{d}}_{t_q}$ collects significant impulse indicators.

Augmented models with forecasted indicators (AF): these models utilize block separation and correct the ragged-edge problem by using forecasts for the missing monthly observations from (3.79) obtained from the blocks of correlated variables. The nowcasting models are then selected by *Autometrics* with IIS recursively using in-sample information on the indicators only. Based on that specification, a one-quarter ahead nowcast is produced by using in-sample observations and forecasted values if the selected model contains lags of leading indicators not yet available. A nowcasting model then has the following form:

$$\Delta\widehat{y}^{h^k_q}_{t_q} = \widehat{\alpha}'\Delta\widehat{\mathbf{y}}^{v}_{t_q} + \widehat{\beta}'\widehat{\mathbf{z}}_{t_q} + \widehat{\gamma}'\widetilde{\mathbf{z}}_{t_q} + \widehat{\zeta}'\widehat{\mathbf{d}}_{t_q} \tag{3.83}$$

where all variables are the same as in (3.82) and $\widetilde{\mathbf{z}}_{t_q}$ are the forecasted values for the selected relevant leading indicators. Comparison between AI and AF should reveal potential benefits from forecasting monthly indicators as compared to nowcasting from in-sample information only.

Augmented models with robust forecasts for indicators (AR): instead of the conditional monthly forecasts from (3.79), a differencing device is used for individual indicators forecasts that is robust after breaks, such that:

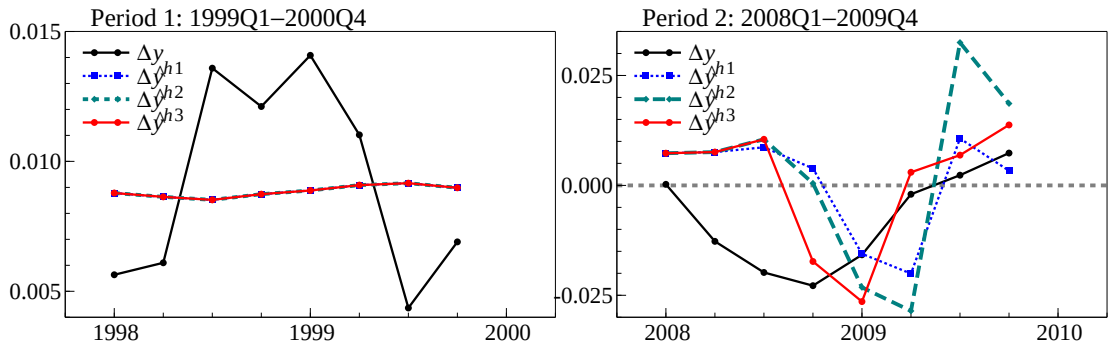
$$\widetilde{z}_{i,t_q} = z_{i,t_q - l_i} \tag{3.84}$$

where l_i is the release lag for variable i. The same type of GUM is used as for AI and AF and the nowcasting equation will include both in-sample and forecasting indicators similarly to (3.83). Note that the selected models for AR and AF will be identical, with the difference in the nowcasted values driven by the methodologies of indicator forecasts.

3.7.5 Vintage models

The nowcasts from VM for the two subsamples are presented in Figure 3.22, which plots the actual GDP growth against the nowcasts at three horizons respectively. For the first subsample, the nowcasts are identical for all three horizons, since the selected models do not differ across the horizons. In fact, *Autometrics* did not select any vintages as relevant and the final specification only includes an intercept, which is the same for all three horizons. Consequently, preliminary vintages do not have any explanatory power for the latest estimate of the GDP growth between 1992 and 2000. The conditional in-sample mean produces the best parsimonious nowcasting model. This is in line with the results obtained in Giannone et al. (2008), who report that during stable growth periods the conditional means are difficult to beat on average.

Figure 3.22: Nowcasts from vintages model for three horizons and two subsamples: 1999Q1–2000Q4 (left panel) and 2008Q1–2009Q4 (right panel).



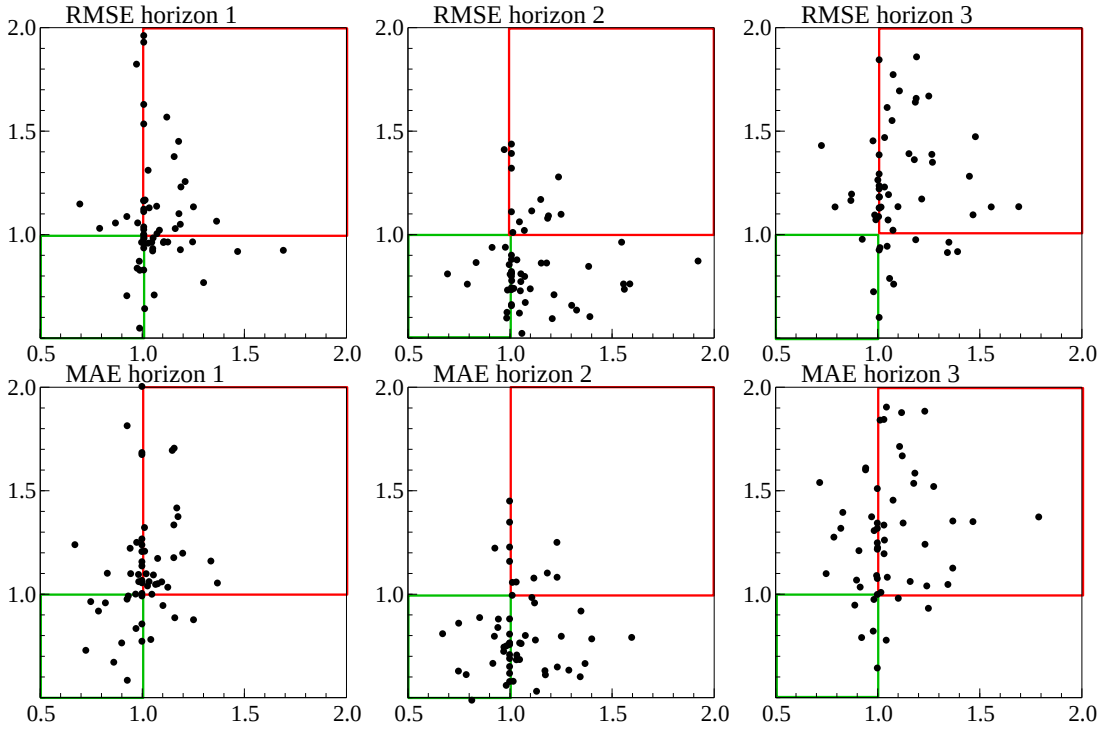
Subsample 2 is characterized by negative growth for 5 consecutive quarters starting from 2008Q2. For this and the subsequent quarter, VM predicts the output to grow. The vintages model nowcasts the start of the recession two quarters after its start, in 2008Q4, and from then on all three vintages can track the series quite closely. This finding, once again, underlines the weakness of the preliminary estimates and their lags in explaining actual growth. At all horizons, the nowcast growth paths are quite similar. Moreover, the model specification at $h^3_{t_q}$, when the first preliminary vintage is released, provides a noticeable improvement for most quarters, as compared to the earlier horizons. However, the first recovery period in 2009Q3 is correctly predicted at all horizons, but the magnitude of the error is quite significant. Still, a promising finding is that the model from horizon 3 that includes $\Delta y^{v_1}_{t_q}$ in the selected specification has the lowest RMSE, which provides some reassurance for applying contemporaneous information to improve the nowcasting accuracy.

3.7.6 Single-indicator models

The single indicator models are compared to the VM model in terms of the relative RMSE and MAE for the two periods. Figure 3.23 depicts the ratios for the single models' RMSEs, to the vintage model's RMSE, and similar ratios for MAEs. Each point corresponds to one single indicator model. The ratio for the first subsample is plotted on the x-axis, and the ratio for the second subsample on the y-axis. If the ratio is smaller than 1, the model performs better than VM for that particular period. A point within the green square corresponds to a model for an indicator that provides consistent improvement relative to VM in both subsamples, whereas a red square corresponds to robustly underperforming indicators.

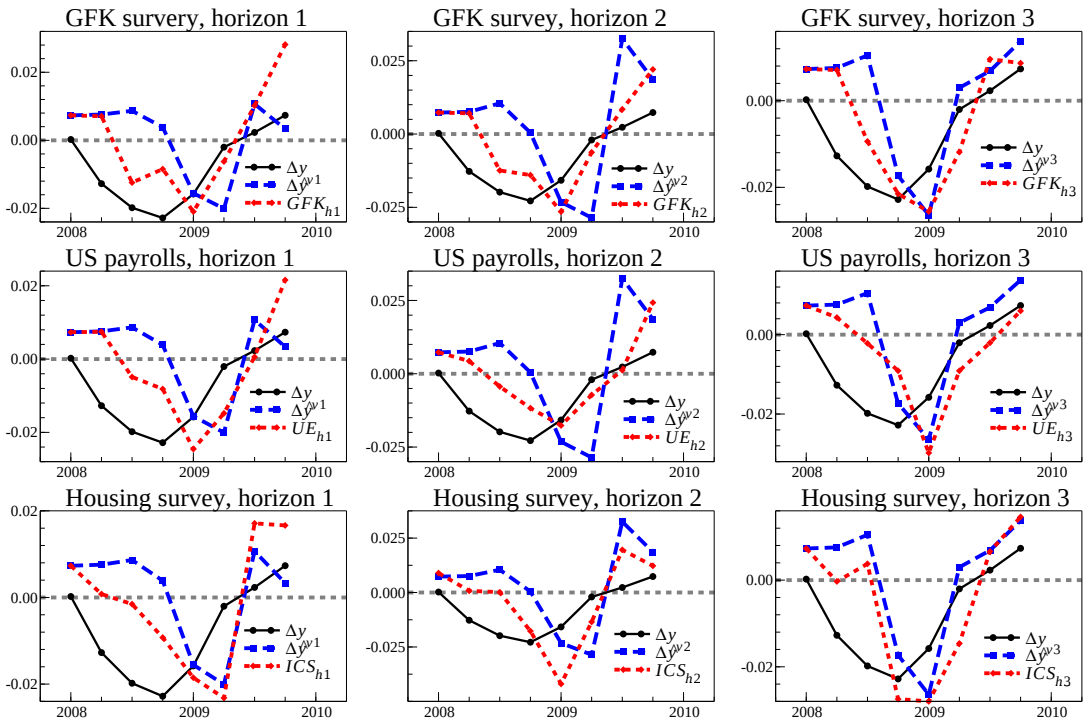
From the pattern emerging in Figure 3.23, there is little evidence of robustly well performing models across the set of indicators. The latter effect would be apparent if the points clustered around the 45 degree line starting from the origin. Only a handful of models provide improvement on VM for every horizon. For instance, at horizon 3 only three indicators perform better in terms of RMSE and MAE. Horizon 2 presents a more promising picture, with around half of the indicators being able to improve on the VM models for the second subsample. Out of these models, 15 are also better predictors of GDP growth in period 1. For horizon 2, the models are least dispersed, but for horizon 3 only a few single models are robustly better than VM.

Figure 3.23: Scatter plots of relative RMSEs (top three plots) and MAEs (bottom three plots) for vintage models augmented by single indicators relative to vintage models for two sub-periods and three nowcasting horizons. Ratios of RMSEs for single models to the vintage models for the first period, 1992–2000, on the x-axis and for the second period, 2001–2009, on the y-axis for the three nowcasting horizons. Red areas define regions of robust underperformance, and green, regions of robust overperformance.



Only three single variables show a robust accuracy improvement across all periods, nowcast horizons and accuracy measures. These are the consumer confidence index, non-farm payrolls and ICS housing survey, all corresponding to block 1. The nowcasting output for the single models corresponding to these three variables together with the VM models for the second subsample are presented in Figure 3.24. Thus, it can be seen *ex post* that these SM models can nowcast the recession one quarter before VM at almost all horizons, while keeping nowcasting errors lower.

Figure 3.24: Comparing VM and SM outcomes.



Overall, there is little indication of consistent model performance for the single models. The number of indicators that overperform VM models varies dramatically across subsamples, nowcasting horizons and nowcasting accuracy measures. In addition, there is no robust evidence that some indicators are completely irrelevant either. Nevertheless, some single indicators can deliver nowcasting improvement for all horizons and subsamples, implying that the chosen information set is not irrelevant and that *Autometrics* might be able to produce more accurate nowcasts when the indicators are combined.

3.7.7 Augmented models

Three versions of the models augmented with all 60 monthly leading indicators are compared to the vintage models. Table 3.5 presents RMSEs and MAEs for the two subsamples and three nowcasting horizons, where the reported values for the augmented models represent the ratios to RMSEs and MAEs of the vintage models, respectively. First, the VM model performs much better on average in period 1, but the reported RMSEs and MAEs are higher between 2001 and 2009. This is expected from the previous results, namely that VM for the first period is essentially estimating the conditional mean, since only a constant is retained in the final models. However, high volatility and the changing signs of the series in the second subsample are not very well captured by the growth estimates alone.

The same outcome for the relative accuracy between the subsamples is found for all augmented models. In addition, a particularly interesting finding for the first subsample is that augmenting the model by leading indicators tends to reduce accuracy on average. It is not the case only for the AG model in horizons 2 and 3, for which the monthly variables slightly improve MAEs of the nowcasts. The models augmented by indicator forecasts perform better than the other two augmented specifications for the later horizons. However, there is no consistent improvement as time elapses, since RMSEs do not reduce monotonically as the nowcast horizon moves closer to the first growth estimate release date.

For some models, nowcasts for horizon 2 are more accurate than those in horizon 3, whereas for others that relationship is reversed. In particular, models with the corrected ragged edges show consistently higher

Table 3.5: RMSEs of augmented models as ratios to RMSEs of the vintage models (top table), MAEs of augmented models as ratios to MAEs of the vintage models (bottom table).

	RMSE					
	1993–2000			2001–2009		
Model	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$
Vintages	*0.004*	*0.004*	*0.004*	*0.017*	*0.022*	*0.014*
Augmented in-sample*	1.178	1.868	1.306	0.691	0.556	0.664
Augmented forecasts*	1.442	1.029	1.142	0.791	0.438	0.689
Augmented robust*	1.476	1.208	1.195	0.784	0.438	0.692

*RMSE ratios to RMSEs of the corresponding vintage models

	MAE					
	1993–2000			2001–2009		
Model	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$	$h^1_{t_q}$	$h^2_{t_q}$	$h^3_{t_q}$
Vintages	*0.004*	*0.004*	*0.004*	*0.014*	*0.020*	*0.011*
Augmented in-sample*	1.014	1.605	1.111	0.686	0.489	0.754
Augmented forecasts*	1.230	0.906	0.996	0.818	0.431	0.788
Augmented robust*	1.294	1.028	1.032	0.793	0.426	0.769

*MAE ratios to MAEs of the corresponding vintage models

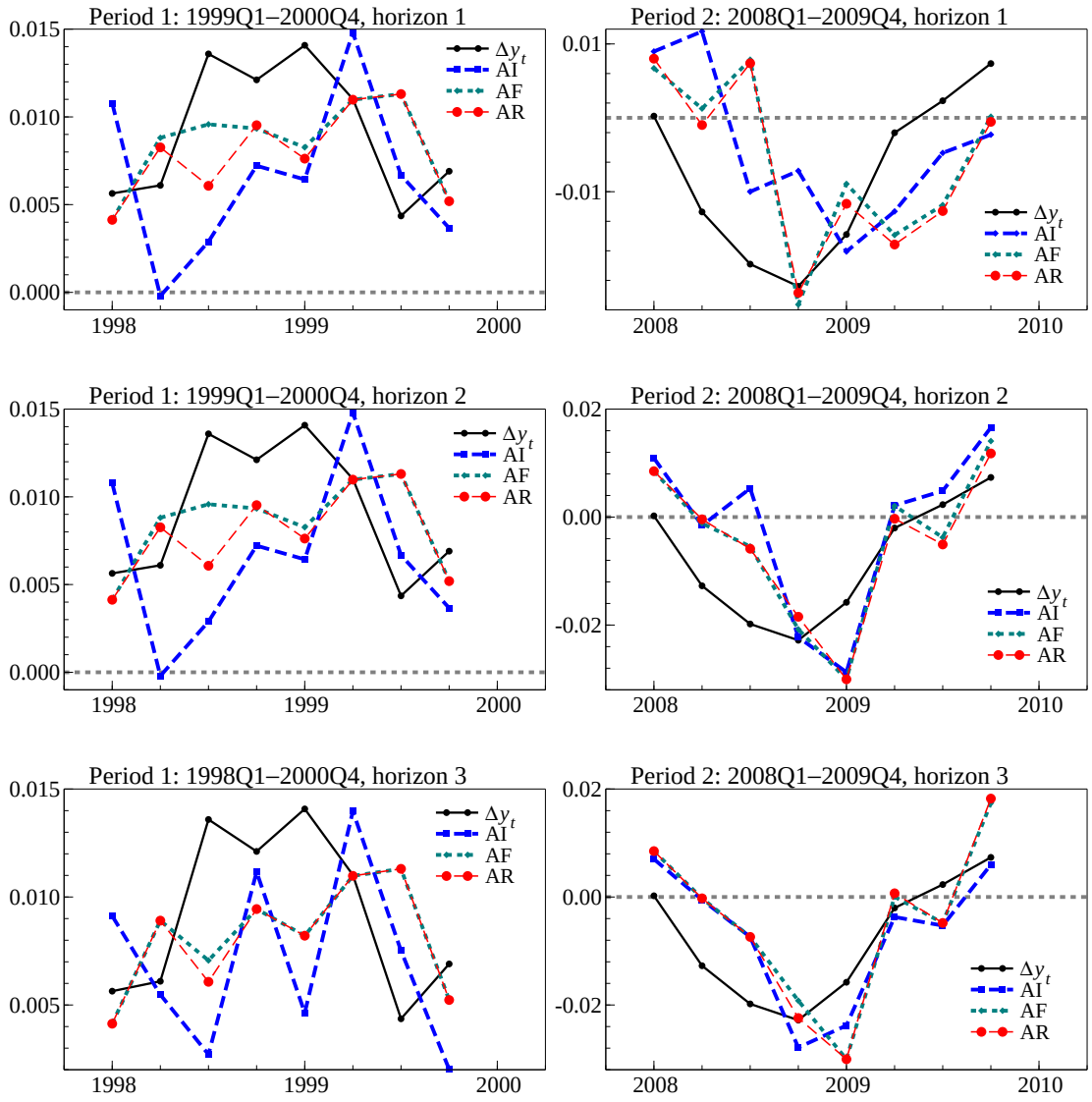
RMSEs and MAEs for the first horizon. This suggests that if the bridge equations framework is used to produce balanced panels by predicting the monthly indicators, the inflow of disaggregated information tends to improve nowcasting accuracy. The opposite is true for the AI model as the most accurate nowcasts are produced in $h^1_{t_q}$. When juxtaposing AI with AR, the latter produce better growth estimates on average when some information about the last month of the quarter becomes available, i.e., horizons 2 and 3. The results for the stable positive growth subsample in general support previous findings for alternative nowcasting methodologies.

In terms of the particular variables selected by *Autometrics* in the final specification of the augmented models for the first subsample, growth rates are never retained. There is some robustness in the selected models for different nowcasting horizons, where the retained variables do not vary considerably among specifications. On average, around 10 lags are selected in each specification. There is much more variability in terms of the selected variables and their lags across nowcasting origins, meaning that there is little robustness across time periods. The variables that were most often retained are the US consumer sentiment index, UK industrial production of various products and the average discount rate for UK Treasury bills. In general, due to lack of consistency in the terminal augmented models and their inferior performance relatively to the vintages model, it is suggested that the leading indicators are not very good predictors of growth when it is stable. The latter effect becomes even more prominent when examining the actual nowcasts produced by the augmented models for the first subsample, depicted in Figure 3.25 for three horizons. Significant differences can be noticed across the models and horizons.

The second subsample represents a more interesting case due to high variability and alternating directions of growth. Table 3.5 shows that unlike in the first subsample, all augmented models provide substantial improvement on the VM, as mirrored by RMSE and MAE ratios much lower than unity. In fact, horizon 2 nowcasts have the greatest relative accuracy enhancement, with both RMSEs and MAEs reduced by a factor larger than 2. At horizons 1 and 3, the ratios are slightly higher. However, horizon 3 has much lower level RMSEs and MAEs for the vintage models induced by inclusion of the first flash contemporaneous growth estimate into the GUM, which is in fact selected by *Autometrics* as relevant at various nowcasting origins. Consequently, it is concluded that in absolute terms, all augmented models produce progressively more accurate nowcasts as new disaggregated information becomes available.

Comparison of the augmented models relative performance reveals that in-sample models perform better for

Figure 3.25: Nowcasts from augmented models for two subsamples: 1999Q1–2000Q4 (left panels) and 2008Q1–2009Q4 (right panels) and three horizons: horizon 1 (top panels), horizon 2 (middle panels), and horizon 3 (bottom panels)



horizon 1 and 3. Although rather insignificant, this finding suggests that using *Autometrics* to forecast individual indicators and then using those forecasts in the nowcasting bridge equations actually reduces overall accuracy. Although true on average, when looking at point nowcasts presented in Figure 3.25, it is apparent that AF and AR can track major turns in the evolution of the GDP growth series slightly better than the AI model. This effect is particularly noticeable in horizons 1 and 2.

One of the key results in this empirical exercise is the satisfactory performance of the robust device as applied to forecasting the leading indicators that face contemporaneous breaks. The model augmented with robust forecasts is, in fact, the one that produces the most timely nowcast of the first instance of output decline. The AR specification correctly predicts negative growth for 2008Q2 at the second month of the quarter, i.e., two months before the release of the first ONS estimate. This finding confirms that a device robust to breaks can yield some improvement in nowcasting accuracy. This provides further support for the results obtained in the simulations above.

Unlike the AR model, AI and AF nowcast the change of the sign in the growth series at the end of the quarter.

During the recessionary period, all models provide relatively good estimates of GDP growth, which are more accurate than the benchmark vintages model and the estimates published by the ONS.

When analyzing the specific variables and their lags retained in final specifications, only one leading indicator is consistently selected as relevant. This is the UK jobless claimants count rate, which is available with a one month delay and is chosen by all models at all horizons and nowcast origins. Various industrial production indices for the UK and the US as well as the GFK consumer confidence index are also chosen in most models. This combination of hard and soft data was expected from their robust performance in the single model specifications and confirms most previous findings. Finally, at all horizons, the financial variables in block 5, such as the FTSE index are occasionally selected. However, the lags vary substantially and the coefficients are almost negligible so have insignificant explanatory power for the growth series. All in all, there is some evidence of consistency in terms of the relevant variables, and very limited indications that the timing of the influence is robust, since the selected lags differ considerably across nowcasting origins and horizons.

3.7.8 Contemporaneous breaks

An examination of the contemporaneous impulse dummies retained in the forecasting models for the monthly indicators is conducted at a nominal 1% significance level and a much tighter 0.1%, suggested as the preferred strategy in Castle, Fawcett, and Hendry (2009). Since the first subsample is characterized by low variance and mean-reverting positive growth, there were no breaks detected for the GDP series at either significance level. Moreover, the monthly indicators were in general quite stable with only a few impulse dummies detained at the 1% level and virtually none at 0.1%. Thus, any further break analysis for the first subsample seems irrelevant.

In the second subsample, on the other hand, multiple breaks are detected for the growth variable. In particular, in-sample breaks are significant at 0.1% level for 2008Q2, 2008Q3 and 2008Q4, the periods of the largest output declines. This confirms that the recession was a location shift relative to previous information for the GDP growth series. Thus, the ability of the augmented nowcasting models to produce good predictions of the output decline further confirms that disaggregated variables contain timely and relevant information to predict major macroeconomic movements, even when the latter have unanticipated and highly significant location shifts.

An analysis of the common contemporaneous breaks was conducted within 6 blocks of correlated variables for the second subsample, and the second quarter of 2008 in particular. At the 1% level, too many breaks were detected to be useful, whereas when a tighter significance level at 0.1% is used, few dummies are retained. However, during the months of 2008Q2, those variables that were retained in the terminal nowcasting models at each horizon do not contain any contemporaneous breaks. On the contrary, variables that do break during these months were not selected. These results provide insufficient evidence for the contemporaneous intercept device to be applied for the correction of the monthly indicators' forecasts. Alternative methods of blocking variables could be examined: testing for cointegration or common serial correlation could reveal links between disaggregated variables.

3.8 Conclusions

First, some conclusions on model-based forecasting in non-stationary economies subject to unanticipated structural breaks, where models differ from DGPs in unknown ways, selected from possibly unreliable data. We showed that the forecasting implications differ considerably from a setting where the model coincided with the DGP in a constant mechanism. That was because unanticipated location shifts are pernicious for forecasting, leading to systematic mis-forecasting in all forms of equilibrium-correction models. Conversely, if no location shift occurred, every DGP parameter could be shifted without any noticeable failure.

That led us to conclude that robust devices could play an invaluable role after location shifts occurred, albeit that they would not forecast such breaks. We explained why the pre-existing conditional expectations, even

given all relevant information, would not provide a good forecast when unanticipated location shifts occurred, which in turn invalidates inter-temporal theory derivations based on the law of iterated expectations. As almost identical location shifts could be created by many different combinations of DGP parameters shifting, and which had changed may not be discernible till well after the event, economic agents as well as nowcasters and forecasters, would be unable to quickly learn how to foresee in the changed environment.

Nevertheless, systematic mis-forecasting could be mitigated by differencing the selected econometric system, retaining original estimates, even if the DGP parameters changed. Devices like $\Delta^2 \widehat{\mathbf{x}}_{T+h|T+h-1} = \mathbf{0}$ are knowingly mis-specified in-sample by restricted information, yet avoid systematic forecast failure on forecasts made after a shift. Moreover, the costs of unnecessary differencing when there is no shift are relatively small compared to the costs of ignoring a break when it has happened. Such results have implications for modeling methodology, as they stress that forecast failure requires a location shift in the DGP relative to the empirical model, and other changes in the DGP may have little observable impact. Thus, the verisimilitude of a model cannot be checked by forecasting success or failure.

Our proposed approach uses *Autometrics* to select over many potentially relevant covariates using higher frequency data, with in-sample and end-point location shift detection based on IIS. Such equations then forecast all disaggregates ready for comparison with incoming measured disaggregates. Forecasts of the missing disaggregates can then be adjusted when there is evidence of location shifts occurring within the period. Alternatively, the robust devices could be used. No device considered here can forecast future location shifts. A different class of model is needed for that, based on different information: see e.g., Castle, Fawcett, and Hendry (2011).

Nowcasting has much in common with, but also differs substantively from, forecasting due to the different roles played by problems of missing data, measurement errors, recording delays, changing databases, and location shifts. The apparent 'break down' of the UK ONS's current methods, noted in the report by the *Financial Times* above, shows the difficulty of nowcasting in times of economic uncertainty and structural change. Location shifts induce nowcast failure, but interact with measurement errors to make discrimination between the causes difficult within the available time horizon for aggregates, although discrimination seems a lesser problem when building an aggregate from large numbers of disaggregates.

The principle findings of the empirical nowcasting exercise for UK quarterly GDP growth confirm that models augmented with disaggregated monthly variables and selected by *Autometrics* with impulse-indicator saturation have superior predictive power to the official estimates and the univariate benchmark models. The relative nowcasting accuracy is further enhanced during the period spanning five consecutive quarters of negative growth, for which highly significant breaks in the series were detected. In addition, it is found that inflow of disaggregated data throughout the reference quarter tends to improve absolute nowcasting accuracy monotonically, irrespective of the information content in the augmented models. On the other hand, the augmented model with contemporaneous break detection, and subsequent robust correction for the leading indicators' forecasts, produces the most accurate nowcasts. In particular, this model predicts the first quarter of output decline in 2008 one month in advance. For comparison, models that use in-sample or forecast conditioning variables without explicitly modeling breaks were able to predict the same event only after the end of the reference quarter. These findings provide substantial evidence that *Autometrics* model selection with IIS can be effectively used for nowcasting GDP from disaggregated data even when contaminated by multiple structural breaks.

Bibliography

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (2010). New EuroCOIN: Tracking economic growth in real time. *Review of Economics and Statistics 92*, 1024–1034.

Andreou, E., E. Ghysels, and A. Kourtellos (2011). Forecasting with mixed-frequency data. See Clements and Hendry (2011), pp. 225–246.

Angelini, E., G. Camba-Méndez, D. Giannone, G. Rünstler, and L. Reichlin (2011). Short-term forecasts of Euro-area GDP growth. *Econometrics Journal 14*, C25–C44.

Angelini, E., J. Henry, and M. Marcellino (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control 30*, 2693–2724.

Artis, M., A. Banerjee, and M. Marcellino (2005). Factor forecasts for the UK. *Journal of Forecasting 24*, 279–298.

Bánbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. See Clements and Hendry (2011), Chapter 7.

Bańbura, M. and G. Rünstler (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*, 333–346.

Banerjee, A., M. Marcellino, and I. Masten (2005). Leading indicators for Euro area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics 67*, 785–813.

Bårdsen, G., A. den Reijer, P. Jonasson, and R. Nymoen (2012). MOSES: Model of Swedish economic studies. *Economic Modelling 29*, 2566–2582.

Bårdsen, G., S. Hurn, and Z. McHugh (2012). Asymmetric unemployment rate dynamics in Australia. Studies in Nonlinear Dynamics & Econometrics, 16 (1).

Barhoumi, K., S. Benk, R. Cristadoro, A. D. Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, C. Van Nieuwenhuyze, and G. Rünstler (2008). Short-term forecasting of GDP using large monthly datasets–A pseudo real-time forecast evaluation exercise. Working paper 133, National Bank of Belgium, Brussels.

Barrell, R. (2001). Forecasting the world economy. In D. F. Hendry and N. R. Ericsson (Eds.), *Understanding Economic Forecasts*, pp. 149–169. Cambridge, Mass.: MIT Press.

Barrell, R., K. Dury, D. Holland, N. Pain, and D. te Velde (1998). Financial market contagion and the effects of the crises in East Asia, Russia and Latin America. *National Institute Economic Review 166*, 57–73.

Boivin, J. and S. Ng (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking 1(3)*.

Bontemps, C. and G. E. Mizon (2008). Encompassing: Concepts and implementation. *Oxford Bulletin of Economics and Statistics 70*, 721–750.

Box, G. E. P. and G. M. Jenkins (1976). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day. First published, 1970.

Brisson, M., B. Campbell, and J. W. Galbraith (2003). Forecasting some low predictability time series using diffusion indices. *Journal of Forecasting 22*, 515–531.

Camacho, M. and I. Sancho (2003). Spanish diffusion indexes. *Spanish Economic Review 5*, 173–203.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). Editors' introduction. In J. Campos, N. R. Ericsson, and D. F. Hendry (Eds.), *Readings on General-to-Specific Modeling*, pp. 1–81. Cheltenham: Edward Elgar.

Castle, J. L., M. P. Clements, and D. F. Hendry (2013). Forecasting by factors, by variables, by both or neither? *Journal of Econometrics 177*, 305–319.

Castle, J. L., M. P. Clements, and D. F. Hendry (2015). Robust approaches to forecasting. *International Journal of Forecasting*, 31(1), 99-112.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics 3 (1)*, DOI: 10.2202/1941–1928.1097.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*, 239–246.

Castle, J. L., J. A. Doornik, and D. F. Hendry (2013). Model selection in equations with many 'small' effects. *Oxford Bulletin of Economics and Statistics 75*, 6–22.

Castle, J. L., J. A. Doornik, D. F. Hendry, and R. Nymoen (2014). Mis-specification testing: Non-invariance of expectations models of inflation. *Econometric Reviews 33*, 553–574.

Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. Econometrics, 3(2), 240-264.

Castle, J. L., N. W. P. Fawcett, and D. F. Hendry (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review 210*, 71–89.

Castle, J. L., N. W. P. Fawcett, and D. F. Hendry (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics 158*, 25–36.

Castle, J. L., N. W. P. Fawcett, and D. F. Hendry (2011). Forecasting Breaks and During Breaks. See Clements and Hendry (2011), pp. 315–353.

Castle, J. L. and D. F. Hendry (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting 29*, 200–214.

Castle, J. L. and D. F. Hendry (2011). Automatic selection of non-linear models. In L. Wang, H. Garnier, and T. Jackman (Eds.), *System Identification, Environmental Modelling and Control*, pp. 229–250. New York: Springer.

Castle, J. L. and D. F. Hendry (2014a). Model selection in under-specified equations with breaks. *Journal of Econometrics 178*, 286–293.

Castle, J. L. and D. F. Hendry (2014b). Semi-automatic non-linear model selection. In N. Haldrup, M. Meitz, and P. Saikkonen (Eds.), *Essays in Nonlinear Time Series Econometrics*, pp. 163–197. Oxford: Oxford University Press.

Castle, J. L. and N. Shephard (Eds.) (2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Choi, H. and H. R. Varian (2012). Predicting the present with Google Trends. *Economic Record 88*, 2–9.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*, 591–605.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed frequency data: Forecasting US output growth. *Journal of Business and Economic Statistics 26*, 546–554.

Clements, M. P. and A. B. Galvão (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics 24*, 1187–1206.

Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P. and D. F. Hendry (2001a). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting 17*, 550–554.

Clements, M. P. and D. F. Hendry (2001b). An historical perspective on forecast errors. *National Institute Economic Review 177*, 100–112.

Clements, M. P. and D. F. Hendry (Eds.) (2002). *A Companion to Economic Forecasting*. Oxford: Blackwells.

Clements, M. P. and D. F. Hendry (2003). Forecasting in the National Accounts at the Office for National Statistics. Report no 12, Statistics Commission.

Clements, M. P. and D. F. Hendry (2005). Guest Editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics 67*, 713–753.

Clements, M. P. and D. F. Hendry (2008). Economic forecasting in a changing world. *Capitalism and Society 3*, 1–18.

Clements, M. P. and D. F. Hendry (Eds.) (2011). *Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Corradi, V., A. Fernandez, and N. R. Swanson (2009). Information in the revision process of real-time datasets. *Journal of Business and Economic Statistics 27*, 455–467.

Croxson, K. and J. J. Reade (2014). Information and efficiency: Goal arrival in soccer betting. *Economic Journal 124*, 62–91.

D'Agostino, A. and D. Giannone (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics 74*, 306–326.

D'Agostino, A., K. McQuinn, and D. O'Brien (2013). Nowcasting Irish GDP. Journal of Business Cycle Measurement and Analysis, OECD, 2012/2.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*, 318–328.

den Reijer, A. (2005). Forecasting Dutch GDP using Large Scale Factor Models. Working paper 028, Netherlands Central Bank, Amsterdam.

Diebold, F. X. and G. D. Rudebusch (1991). Turning point prediction with the composite leading index: An ex ante analysis. In K. Lahiri and G. H. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records*, pp. 231–256. Cambridge: Cambridge University Press.

Diron, M. (2008). Short-term forecasts of Euro-area real GDP growth: An assessment of real-time performance based on vintage data. *Journal of Forecasting 27*, 371–390.

Doornik, J. A. (2007). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics 70*, 915–925.

Doornik, J. A. (2009a). Autometrics. See Castle and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Improving the timeliness of data on influenza-like illnesses using Google Trends. http://www.doornik.com/flu.

Doornik, J. A. and H. Hansen (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics 70*, 927–939.

Doornik, J. A. and D. F. Hendry (2013). *Empirical Econometric Modelling using PcGive: Volume I* (7th ed.). London: Timberlake Consultants Press.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics 164*, 188–205.

Eickmeier, S. and C. Ziegler (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting 27*, 237–265.

Elliott, G., C. W. J. Granger, and A. Timmermann (Eds.) (2006). *Handbook of Econometrics on Forecasting*. Amsterdam: Elsevier.

Emerson, R. A. and D. F. Hendry (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting 15*, 271–91.

Engle, R. F. and D. F. Hendry (1993). Testing super exogeneity and invariance in regression models. *Journal of Econometrics 56*, 119–139.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Ericsson, N. R. (2012). Detecting crises, jumps, and changes in regime. Working paper, Federal Reserve Board of Governors, Washington, D.C.

Espasa, A. and I. Mayo-Burgos (2012). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting 29*, 718–732.

Faust, J., J. H. Rogers, and J. H. Wright (2007). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking 37*, 403–419.

Ferrara, L., D. Guegan, and P. Rakotomarolahy (2010). GDP nowcasting with ragged-edge data: A semi-parametric modelling. *Journal of Forecasting 29*, 186–199.

Fildes, R. and K. Ord (2002). Forecasting competitions–their role in improving forecasting practice and research. See Clements and Hendry (2002), pp. 322–253.

Fildes, R. A. and S. Makridakis (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review 63*, 289–308.

Fixler, D. J. and B. T. Grimm (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business 88*, 16–32.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2001). Coincident and leading indicators for the Euro Area. *Economic Journal 111*, 62–85.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). Do financial variables help forecasting inflation and real activity in the Euro Area? *Journal of Monetary Economics 50*, 1243–1255.

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*, 830–840.

Frisch, R. (1938). Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in Hendry D. F. and Morgan M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Frisch, R. and F. V. Waugh (1933). Partial time regression as compared with individual trends. *Econometrica 1*, 221–223.

Gertler, M. and K. Rogoff (Eds.) (2005). *NBER Macroeconomics Annual 2004*. Cambridge Mass.: MIT Press.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: MIxed DAta Sampling regression models. mimeo, Chapel Hill, N.C.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews 26*, 53–90.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 1545–1578.

Giannone, D., L. Reichlin, and S. Simonelli (2009). Nowcasting Euro Area economic activity in Real-Time: The role of confidence indicators. *National Institute Economic Review 210*, 90–97.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting GDP and Inflation: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*, 665–676.

Gil, R. and S. D. Levitt (2007). Testing the efficiency of markets in the 2002 World Cup. *Journal of Prediction Markets 1*, 255–270.

Godfrey, L. G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables *Econometrica 46*, 1303–1313.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica 12*, 1–118. Supplement.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural Change and Economic Dynamics 11*, 45–65.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics 135*, 399–426.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T. C. Mills and K. D. Patterson (Eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *Econometrics Journal 7*, 1–31.

Hendry, D. F. and J. A. Doornik (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge, Mass.: MIT Press.

Hendry, D. F. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of Business and Economic Statistics 29*, 216–227.

Hendry, D. F. and S. Johansen (2015). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, 31, 93-114.

Hendry, D. F., S. Johansen, and C. Santos (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 33*, 317–335. Erratum, 337–339.

Hendry, D. F. and H.-M. Krolzig (2003). New developments in automatic general-to-specific modelling. In B. P. Stigum (Ed.), *Econometrics and the Philosophy of Economics*, pp. 379–419. Princeton: Princeton University Press.

Hendry, D. F. and H.-M. Krolzig (2005). The properties of automatic Gets modelling. *Economic Journal 115*, C32–C61.

Hendry, D. F. and G. E. Mizon (2011a). Econometric modelling of time series with outlying observations. *Journal of Time Series Econometrics 3 (1)*, DOI: 10.2202/1941–1928.1100.

Hendry, D. F. and G. E. Mizon (2011b). What needs re-thinking in macroeconomics? *Global Policy 2*, 176–183.

Hendry, D. F. and G. E. Mizon (2012a). Forecasting from structural econometric models. Working paper 597, Economics Department, Oxford University.

Hendry, D. F. and G. E. Mizon (2012b). Open-model forecast-error taxonomies. In X. Chen and N. R. Swanson (Eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

Hendry, D. F. and G. E. Mizon (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics 182*, 186–195.

Hendry, D. F. and F. Pretis (2013). Anthropogenic influences on atmospheric CO2. In R. Fouquet (Ed.), *Handbook on Energy and Climate Change*, pp. 287–326. Cheltenham: Edward Elgar.

Hendry, D. F. and J. J. Reade (2008a). Economic forecasting and prediction markets. Working paper, Economics Department, Oxford University.

Hendry, D. F. and J. J. Reade (2008b). Modelling and forecasting using model averaging. Working paper, Economics Department, Oxford University.

Hendry, D. F. and J.-F. Richard (1989). Recent developments in the theory of encompassing. In B. Cornet and H. Tulkens (Eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press.

Hendry, D. F. and C. Santos (2010). An automatic test of super exogeneity. In M. W. Watson, T. Bollerslev, and J. Russell (Eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford: Oxford University Press.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Hubrich, K. (2005). Forecasting Euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting 21*, 119–136.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. See Castle and Shephard (2009), pp. 1–36.

Kapetanios, G. and M. Marcellino (2010). Factor-GMM estimation with large sets of possibly weak instruments. *Computational Statistics and Data Analysis 54*, 2655–2675.

Kitchen, J. and R. M. Monaco (2003). Real-time forecasting in practice: the U.S. Treasury staff's real-time GDP forecast system. *Business Economics October*, 10–19.

Kitov, O. I. (2012). Nowcasting GDP from disaggregated data in the presence of structural breaks. Unpublished M.Phil. Thesis, University of Oxford.

Kuzin, V., M. Marcellino, and C. Schumacher (2009). Pooling versus model selection for nowcasting with many predictors: An application to German GDP. Discussion paper series 1: Economic studies, no. 3, Deutsche Bundesbank.

Kuzin, V., M. Marcellino, and C. Schumacher (2012). Pooling versus model selection for nowcasting with many predictors: Empirical evidence for six industrial countries. Journal of Applied Econometrics, 28(3), 392-411.

Lee, K., N. Olekalns, and K. Shields (2008). Nowcasting, business cycle dating and the interpretation of new information when real time data are available. Discussion papers in economics 08/17, University of Leicester.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. In K. Brunner and A. Meltzer (Eds.), *The Phillips Curve and Labor Markets*, Volume 1 of *Carnegie-Rochester Conferences on Public Policy*, pp. 19–46. Amsterdam: North-Holland Publishing Company.

Lütkepohl, H. (2006). Forecasting with VARMA models. See Elliott et al. (2006), pp. 288–325.

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting 1*, 111–153.

Makridakis, S. and M. Hibon (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting 16*, 451–476.

Marcellino, M. and C. Schumacher (2010). Factor-MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics 72*, 518–550.

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

Mariscal, R. and A. Powell (2014). Commodity price booms and breaks: Detection, magnitude and implications for developing countries. Discussion paper, Research Department, Inter American Development Bank.

Mills, T. C. (2010). Bradford Smith: An econometrician decades ahead of his time. *Oxford Bulletin of Economics and Statistics 73*, 276–285.

Mitchell, J. (2009). Where are we now? The UK recession and nowcasting GDP growth using statistical models. *National Institute Economic Review 209*, 60–69.

Mitchell, J., R. J. Smith, M. R. Weale, S. Wright, and E. Salazar (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth. *Economic Journal 115*, F108–F129.

Mizon, G. E. and J.-F. Richard (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica 54*, 657–678.

Montgomery, A. L., V. Zarnowitz, R. S. Tsay, and G. C. Tiao (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association 93*, 478–493.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least squares regression analysis *Journal of the Royal Statistical Society B 31*, 350–371.

Rapach, D. E. and M. E. Wohar (Eds.) (2008). *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Bingley, UK: Emerald Group.

Reed, G. (2000). How the preliminary estimate of GDP is produced. *Economic Trends 556*, 53–61.

Reed, G. (2002). How much information is in the UK preliminary estimate of GDP? *Economic Trends 585*, 1–8.

Rünstler, G. and F. Sédillot (2003). Short-term estimates of Euro area real GDP by means of monthly data. Working Paper series, no. 276, European Central Bank.

Salkever, D. S. (1976). The use of dummy variables to compute predictions, prediction errors and confidence intervals. *Journal of Econometrics 4*, 393–397.

Schumacher, C. and J. Breitung (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting 26*, 271–302.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*, 386–398.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–464.

Smith, B. B. (1927). Forecasting the volume and value of the cotton crop. *Journal of the American Statistical Association 22*, 442–459.

Smith, B. B. (1929). Judging the forecast for 1929. *Journal of the American Statistical Association 24*, 94–98.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*, 11–30.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indices. *Journal of Business and Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. See Elliott et al. (2006), pp. 515–554.

Stone, J. R. N. (1947). On the interdependence of blocks of transactions. *Journal of the Royal Statistical Society 8*, 1–32. Supplement.

Wallis, K. F. (1986). Forecasting with an econometric model: The 'ragged edge' problem. *Journal of Forecasting 5*, 1–14.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity *Econometrica 48*, 817–838.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science 6*, 324–342.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *Journal of Economic Perspectives 18*, 107–126.

4 The Trade-Off Between Timeliness and Accuracy: The Perspective of a Statistical Agency

Authors

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

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

4.1 Introduction

Statistics are fundamentals in our daily lives; we are surrounded by information about all kind of events, which is more and more easily available. Statistics can be produced by a number of bodies and can have different degrees of trustability. Official statistics do have a privileged position with respect to all other statistical information in this context, in fact they are subject to a strong quality control and calculated following a series of rigorous principles. In the case of European statistics, as an example, they are subject to a continuous evaluation process based on a quality framework composed of seven quality dimension (QFA for ESS , 2011).

Furthermore the compilation of European statistics must follow a series of ethic and behavioural principles collected in the so-called Code of Practice (ESS Code of practice , 2011). Looking at the quality framework, it appears clear that optimising several quality dimensions at the same time is not always possible; indeed the relationship among quality components can be of such a nature that it would be impossible to maximise them simultaneously. Probably the most known case is represented by the so called trade-off between timeliness and accuracy.

Statistics are linked to the concept of measurement of some phenomenon: economic, social, socio-economic, etc.. Trustable measures of phenomena should be as much as possible accurate, meaning reliable and precise; at the same time their usefulness increases often with their timeliness; statistical agencies are then often confronted with the problem of finding the right balance between timeliness and accuracy of their statistics.

In this chapter we aim to point out to some issues involved in the trade-off between timeliness and accuracy. Most official statistics and macroeconomic ones in particular, deal with the measurement of complex phenomena for which strong fundamentals are essential. A primary consideration is that the target variable to be measured has to follow a set of definitions, concepts, methods and standards agreed at international level and evolving over the time. This has to be kept well in mind whenever assessing the accuracy of a measure. International standards define in detail what is to be measured and, in order to keep up to date with an evolving society, have to be regularly revised and improved to better define phenomena under measurement. An example is the system of national accounts which is based on the NSA and the ESA at European level, ever-evolving classifications.

This aspect is linked to the concepts of accuracy and reliability; measuring GDP cannot be linked to the "true value" of this economic variable as it would be the case of a physical measure such as a length; when a producer wishes to assess the accuracy of a GDP estimate, he will look at how much this estimate is revised and how much it is in line with the different GDP definitions. A source of errors in the measurement is the completeness of source data. When data are based on surveys there will be errors due to sampling; moreover, whenever source data are incomplete, because data collection can be complex in some cases, modelling and estimation techniques can be a further cause of errors and impact the estimate accuracy. As a consequence, we could imagine that waiting longer after the end of the reference period should improve data coverage and then accuracy.

What about the user? In the context of macroeconomic variables, users can be policy makers, central bankers or economic and financial market analysts which need to take decisions or get the results of their analyses in a short period, so that their priority would be more on the timeliness, that is data as soon as possible, than on accuracy; users however need also to trust data, a combination of those two aspects is then required, and this is what is called the trade-off between timeliness and accuracy. In this chapter we analyse in details this trade-off with particular focus on its interpretation and possible limitations.

The structure of the chapter is as follow: section 4.2 will introduce background information with a particular focus on the progress in Europe with the introduction of some flash estimates which have only partially mitigated the gap in timeliness with US statistics; section 4.3 will details some aspects of the timeliness accuracy trade-off in the context of policy making; section 4.4 will shortly frame this trade-off in the context of business cycle phases; section 4.5 will introduce some brief considerations related to big data, and section 4.6 will conclude.

4.2 Historical background

The trade-off between timeliness and accuracy has been always a main question for policy makers, in particular in the context of macro-economics. The choice of the best balance balance between timeliness and accuracy is strongly related to the characteristics of the statistical system of each country as well as to the users' requests. This explains why relevant differences in timeliness are observed across countries and regions. As an example, historically, the US have put a lot of emphasis on timeliness, also thanks to a very efficient and centralised statistical system. All macro-economic US variables have been available very timely, and sometime released even before the end of the reference period as it was the case for the so called GNP flash estimates published since 1983 at t-15 days. This sort of flash estimates, previously computed for several years for restricted governmental users only, was subject to criticism from the media which raised doubts on its accuracy, especially when looking at its revisions induced by subsequent national accounts releases.

As stated in Fixler (2007), despite the fact that an analysis of the revisions process did not show any particular strange behaviour, in January 1986 the US BEA decided to discontinue the t-15 estimates. Then the US decided to move towards a more conservative estimate at t+30 days after the end of the reference period. This is almost a unique example of moving toward less timely data in order to avoid users' confusion.

The situation in the European Union (EU) has evolved in a quite different way; in Europe the focus has been more on accuracy at least until the end of the 80s. The amount of infra-annual information collected at the EU level was at the time quite limited, also because there were no specific need for high frequency data for policy making. Furthermore, several statistical offices did not compile at all infra-annual statistics such as quarterly national accounts. The situation started to change with the common market, the Maastricht treaty and especially with the monetary union. The adoption of the euro and the development of a common monetary policy required reliable and timely available infra-annual macro-economic statistics.

For this reasons several initiatives were undertaken jointly by the European Commission and the European Central Bank. The first one was the creation of the so-called Benchmarking task force mandated to investigate the timeliness differentials within the European Union and with the US and Japan, and to understand their main causes. The results of this timeliness comparison as in 2002 is presented in table 4.1 in line with Öberg (2002).

The second initiative was to create the Friends of the Chair Group in order to promote the timeliness and further harmonisation of infra-annual European economic statistics; in this context the Group identified a set of nineteen Principal European Economic Indicators (PEEIs), to be subject to regular monitoring in terms of timeliness, accuracy and coverage. Such a list was officially approved and published in the communication of the Commission to the European parliament and the Council on Eurozone statistics: "Towards improved methodologies for Eurozone statistics and indicators" (COMM 2002/667). The main results of the first investigations of the Benchmarking task force first, and of the Friends of the Chair Group then, were that the European Statistical System is very different from the two others analysed and especially from the US one, mainly in terms of the degree of decentralisation.

A higher decentralisation makes more difficult the undertaking of measures to speed up the production process and data transmission. For this reason, the emphasis in improving timeliness was put on the use of new statistical techniques, such as the construction of flash estimates, or the use of partially incomplete sampling techniques at national level. Thanks to those initiatives, some relevant improvements have been made at the euro area and European Union level even if gaps with the US still persist, especially in some areas such as labour market and short-term business statistics. Table 4.2 presents the timeliness of the PEEIs for the euro area and the European Union as currently observed, and the short-term target for the euro area.

Table 4.1: Principal European Economic Indicators list: Comparison of EU/EMU and US delays (in calendar days)

Set	Indicator	EU delay target	EU delay actual	US delay
Set 1: Price Indicators				
1.1.	Harmonised Consumer Price Index: MUICP flash estimate		2	NA
1.2.	Harmonised Consumer Price Index: Actual indices	17	17	16
Set 2: National Accounts Indicators				
2.1.	Quarterly National Accounts: Flash GDP	30/45	NA	30
2.2.	Quarterly National Accounts: First GDP release with breakdowns	60/90	70/120	30
2.3.	Quarterly National Accounts: Sector Accounts	90	NA	NA
2.4.	Quarterly Government Finance Statistics	90	80 (still annual)	NA
Set 3: Business Indicators				
3.1.	Industrial production index	30	48	16
3.2.	Industrial output price index for domestic markets	20	33	11
3.3.	Industrial new orders index	30/50	soon	26
3.4.	Industrial import price index	30	under develop-ment	15
3.5.	Production in construction	30/45	75	30
3.6.	Turnover index for retail trade and repair	30	60	15
3.7.	Turnover index for other services	30	soon	NA
3.8.	Corporate output price index for services	30	under develop-ment	NA
Set 4: Labour Market Indicators				
4.1.	Unemployment rate	30	30	5
4.2.	Job vacancy rate[a]	30	NA	NA 30
4.3.	Employment index	30	70/75	5
4.4.	Labour cost index Employment cost index	60	90	25
Set 5: Foreign Trade Indicators				
5.1.	External trade balance: intra- and extra-MU; intra- and extra-EU	45	50	48

Source: Öberg (2002), Annex 1B to CPS 2002/45/8.

[a]The help wanted index in the US comes close to providing the type of information contained in a job vacancy rate

Table 4.2: Principal European Economic Indicators: Euro area timeliness and short-term target and EU timeliness

Set	Indicator	Timeliness euro area	Short-term target euro area	Timeliness EU
Set 1: Price Indicators				
1.1.	Harmonised Consumer Price Index: Euro area flash estimate	0	0	-
1.2.	Harmonised Consumer Price Index	17	17	17
Set 2: National Accounts Indicators				
2.1.	Quarterly National Accounts: Flash/first GDP	31/45	30/45	31/45
2.2.	Quarterly National Accounts: GDP estimate and breakdowns	66	60	66
2.3.	Quarterly National Accounts: Household and company accounts	102	90	120
2.4.	Government Finance Statistics	115	90	115
Set 3: Business Indicators[a]				
3.1.	Industrial production	42	40	42
3.2.	Industrial producer prices	33	35	33
3.3.	Industrial import prices	35	45	-
3.4.	Production in construction	47	45	47
3.5.	Retail trade turnover	34	30	34
3.6.	Services turnover	62	60	62
3.7.	Services producer prices	93	90	93
Set 4: Labour Market Indicators				
4.1.	Unemployment rate	30	30	30
4.2.1.	Job vacancy rate: Flash estimate	50	45	50
4.2.2.	Job vacancy rate	78	75	78
4.3.	Employment	74	45	74
4.4.	Labour cost index	78	70	78
Set 5: Foreign Trade Indicators				
5.1.	International trade in goods balance	46	46	46

Source: EFC Status Report, 2017

[a] Industrial new orders has been discontinued.

4.3 Timeliness and policy: The trade-off in the context of policy making

In this section we enlarge the discussion on the trade-off between timeliness and accuracy; we already mentioned that producers tend to maximise accuracy, while users can in some cases be keener of timelier data; however, looking at those two dimensions out of a larger context can be misleading. It would be more interesting to broaden the analysis to some other dimensions: the relevance of the statistics in question, the level of disaggregation needed and the resources necessary to produce them. As observed in Bier and Ahnert (2001), the more important data are for policy decisions and reasoning, the higher is the reliability requirement. Moreover, as Öberg (2002) wrote: "*. . . given limited financial resources, you can always ask what is most important to improve, the accuracy or the timeliness of present short-term economic statistics. This is an issue that is not possible to answer from a statistical standpoint only. Rather, it is an issue of what the users view as most important.*"

Statistics are at the service of the public community and policy makers, but statistical offices and more in general statistical producers rely on a limited amount of resources. With unlimited resources, new surveys or new collection methods could be easily implemented and timeliness decisively improved. Then the first trade-off in front of producers is between available resources and timeliness. A second point to consider, which is particularly important for the users, is the relevance of the statistics and the needed level of disaggregation, where disaggregation could concern geographical level or disaggregation by branches or sectors, as for example when statistics are produced according to the NACE or the COICOP or another classification. Some statistics mighty be highly relevant in the context of policy making, even at detailed level; while for some other ones the relevance of very timely data would be considered as a minor problem.

Statistical producers have to find the right balance among all those aspects, and define clear, and transparent for the users, priorities. Users' needs, of policy makers but not only, and producers' needs have to be balanced in this wider context covering timeliness, resources, and relevance of statistics. The relation among those aspects would not be a linear one; investment in resources would not always improve timeliness in the same proportion, and the search for an optimal and concretely realisable compromise is of outmost importance.

Moreover, in specific contexts, such as highly relevant macroeconomic indicators, the user could more easily accept a minor loss in accuracy in return for improved timeliness. This relation too will not be a linear one, and could depend upon the user profile. Some studies have also shown that improving timeliness does not always imply a loss in accuracy because improvement can be reached thanks to gain in efficiency in the production process (e.g. Fixler (2007)).

All in all, the statistical producer will have to first analyse the priority linked to the timeliness of the statistics, then assign the correct resources to its production, and then consider that a certain increase in the assigned resources will correspond to an increase in timeliness but in a non-linear way. At the same time he should consider what the user satisfaction will be with respect to the producer choice, and eventually adapt efforts in resources and/or methodological improvements to find the right balance between users' needs and the production of reliable and fit for purpose statistics. And the context could be even more complex, as observed by Fixler (2007): "*The magnitude of the trade-off, however, has a subjective component. Different users have different preferences for timeliness and, most important, they have different preferences for the optimal combination of accuracy and timeliness. The needs of a central bank policy staff are different from those of business decision-makers. Statistical agencies must balance the preferences of all of their data users* ".

There is another relevant aspect to consider: the assessment of results; how to assess if an improvement in timeliness has reduced or not the accuracy of our estimates? We can assess the estimates' volatility and their revisions; however, an estimate that is released earlier can be compared with the following estimate, usually released some weeks or days after, or to a final estimate, which can be available after years; results can considerably differ and it is important to have a clear and transparent position on this issue. Usually, policy makers are more interested in a comparison with the following estimate, in order to be able to anticipate their

political choices, and much less keen of information as available after a long while.

Another aspect which can be of some relevance at the level of international organisations is the legal constraints and the different publication practices of member countries (see Bier and Ahnert (2001)), which could be a big obstacle to the improvement of timeliness. It would be quite difficult to improve the timeliness of a geographical aggregate when the legal requirements of the different member countries considerably differ; this is where the need for harmonisation practices and/or legal acts at supranational level can be of value. At European level, transmission of data following precise deadlines is often governed by European legal acts with constraining power on European Member States; continuous efforts towards the harmonisation of release and revisions policy is undergoing, with evident positive impact on the improvement of timeliness; the examples of some countries and the investment in resources at European level has often paved the way to general improvement in timeliness without any loss in accuracy.

This handbook presents techniques to bypass, in particular cases, the limitations due to data collection and production by the application of econometric and modelling techniques. Thanks to those techniques, the equilibrium point between producers' and users' needs can be moved farther towards the users, increasing timeliness without relevant losses of accuracy. Those techniques are a powerful instrument in the hand of statisticians, who in areas where timeliness is particularly important, should systematically produce preliminary statistics and flash estimates as a complement to later published final statistics (Öberg (2002)).

Moreover, official statisticians are in principle warrantors of quality and transparency; when early estimates are produced in the context of official statistics, methods should be as much as possible well known and documented and processes replicable; as a consequence the user should be able to trust results. Early estimates should be published with relevant documentation and the producers should answer specific users' questions. In this context, official statistics can be brought closer to users' needs.

At the end of this section some considerations arise from the previous detailed investigation of the trade-off between timeliness and accuracy. The first one is that this trade-off is not a clear relationship between two quality dimensions but a phenomenon which, under certain conditions, can materialize while under other conditions would not. As already observed by Öberg (2002), if the timeliness is improved thank to modernisation or rationalisation of the production process allowing for a more efficient processing of the information in a reduced number of days, then timeliness will be enhanced without any loss of accuracy since the information set remain unchanged. By contrast it may even happen that the rationalisation of the production process brings an improvement of the accuracy too.

Another consideration is that the acquisition of information, especially at the aggregated level, is usually not a time-continuous process because the information become available at well specified points in time after the reference period. In this situation it is possible to anticipate a data release, and then timeliness, without any loss in accuracy between two specific points in time if no new data enter the information set in-between. For example in Mazzi, Mitchell and Montana (2014) and Mazzi, Moauro and Ruggeri Cannata (2016), when constructing GDP estimates using econometric techniques, it was noted that between 15 and 30 days after the end of the reference period no relevant information was released, so that estimating the GDP at t+15 and at t+30 was producing exactly the same results.

The third consideration is that when working with incomplete information sets to produce rapid estimates, a possible consequence could be a loss in accuracy. Nevertheless this loss could be, at least partially, mitigated by the use of particularly effective modelling techniques designed to maximize the use of the available information.

Finally we would like to focus on an aspect of primary importance, at least in our opinion, especially for developing countries. Identifying the right balance between timeliness and accuracy is a decision which makes sense only if the level of accuracy of the statistics produced by a certain statistical institution is high enough to ensure that a small reduction would not compromise the trustability and credibility of data. When the accuracy of statistics is not particularly high, deciding to further reduce it in favour of a higher timeliness can have very negative consequences not only in terms or credibility of the institution but also for the users.

Providing users with timely but highly imprecise and maybe misleading data could compromise their analysis and decisions, with very negative consequences for their businesses and eventually for the economic situation of the country.

4.4 Timeliness, accuracy and the business cycle phases

Looking at the historical evolution of a given indicator, it is possible to observe that the timeliness is usually constant over a sufficiently long time interval until new decisions aiming to improve it are taken. By contrast, over the same time interval the accuracy of the same indicator is not necessarily constant.

It is possible to show that the accuracy can vary according to cyclical phases, and it tends to deteriorate when approaching turning points. This might be due to the higher degree of uncertainty characterising economic agents which is negatively reflected in the source data used to compile the indicator. Furthermore, other data treatments such as missing data imputation, seasonal adjustment, etc. can be highly sensitive to data uncertainty. The result is that in proximity of turning points data tend to be more volatile, and consequently further revised, with revisions' size usually bigger than in normal periods; this affects data reliability which is a component of the accuracy. By consequence, when evaluating the effect of an increase in timeliness in terms of a potential loss in accuracy, is better to conduct such evaluation over a period not affected by relevant cyclical phenomena. Furthermore, it would be beneficial to carry out this analysis on the base of historical vintages, that is using data as they were published at a certain moment in time, instead of using final vintages, that is the current version of data.

4.5 Big data and the trade-off between timeliness and accuracy

One of the key innovative facts with enormous potential for the future having characterised the beginning of the 21st century is the so-called data revolution. In this context an enormous amount of data coming from different sources such as mobile phone networks, internet, social media, sensors, electronic payment systems etc. have been progressively made available. They could constitute a still largely unexplored information world, which could in future impact the production of official statistics.

Big data have recently attracted the attention of official statisticians and statistical offices are more and more exploring the use of big data sources for the production of official statistics. Using big data remains very challenging for several reasons; the first one is represented by their regular availability over the time since there is no guarantee that data available today will be available in future too. The second is represented by the fact that big data are generally unstructured and they need to be converted into structured panel or time series data. The third one is related to the presence of a large component of noise and to high frequency fluctuations, they need then to be filtered in order to extract a proper signal. The literature on this topic is quite large; in this short section we just shortly mention some considerations on how big data could impact the trade-off between timeliness and accuracy.

Especially in the macro-economic domain, the importance of big data has been tested mainly in the construction of nowcasting and other kind of rapid estimates. Chapter sixteen will discuss those aspects in several details.

Looking at big data, it appears quite clear that the almost real-time data collection process makes data available just at the end of the reference period so that timeliness should not be in question anymore.

Nevertheless, results obtained in building up rapid estimates using big data sources only, are at present very encouraging but not yet conclusive. By contrast, big data can be used to complement traditional sources of

information in building up new kind of rapid estimates, either to increase timeliness or to reduce the accuracy loss at a given timeliness. This is a very promising area which however remains at present mostly a research topic.

4.6 Conclusions

In this chapter we have investigated several aspects related to the so-called trade-off between timeliness and accuracy. In particular, we have focused our attention on how this trade-off has been historically interpreted across countries and regions, especially US and Europe. Then, we tried to give an interpretation of this trade-off in the light of a rational statistical office aiming to satisfy users' needs and production constraints in a given limited resources context. Starting from this analysis, we have investigated special cases were the trade-off between timeliness and accuracy does not hold; in this context we have also shown as under special conditions timeliness and accuracy might even be simultaneously improved. The fact that the improvement of timeliness is an objective to be achieved only when data are accurate and trustable enough has been mentioned, with particular reference to developing countries. Finally, we have shortly investigated the impact on the trade-off between timeliness and accuracy of the economic cyclical situation and of the use of big data as a complementary data source in the construction of rapid estimates.

Bibliography

W. Bier, H. Ahnert (2001), "Trade-off between timeliness and accuracy, ECB requirements for general economic statistics", Article published in Dutch in "Economisch Statistische Berichten (ESB)" No. 4299.

Communication of the Commission to the European Parliament and the council on Eurozone statistics: "Towards improved methodologies for Eurozone statistics and indicators", COMM/202/661, available on-line at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0661:FIN:EN:PDF.

European Statistics Code of Practice - revised edition 2011, available on-line at: http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF

D. Fixler (2007), Timeliness and Accuracy, Measurement in Economics, A Handbook edited by Marcel Boumans, Academic Press, London.

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, available on-line at: http://ec.europa.eu/eurostat/documents/3888793/7230975/KS-TC-16-002-EN-N.pdf.

S. Öberg (2002), The future of the European Statistical System, 88th DGINS Conference, Palermo 19-20 September 2002.

Quality Assurance Framework of the European Statistical System, available online at: http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf.

11 Statistical and Econometric Techniques for Rapid Estimates

5 An Overview of Modelling Techniques for Rapid Estimates

Authors

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Riccardo Gatto
Eurostat
riccardo.gatto@ec.europa.eu

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

5.1 Introduction

In recent years, the need for improving timeliness mainly, but not exclusively, of infra-annual macroeconomic statistics became a priority at both national and international level. Policy and decision makers, and in particular the management of a monetary policy for the Euro area, require readily-available macroeconomic information to design and implement effective monetary and economic policy measures. Such a situation is well described by the situation of European statistics which are supposed to support common monetary policy decisions at the Euro area level as well as EU and national economic policies. At the EU level, the timeliness of key macroeconomic indicators, with a few exceptions, is lower than in some non-European countries. Concerning timeliness, the US has historically represented the best example to be followed by statistical systems all around the world. Several studies and initiatives have been undertaken to improve timeliness of infra-annual macroeconomic statistics leading to significant improvements at the European level but also for many other countries such as China. Unfortunately, the 2007-2009 global financial and economic crisis has shown that the timeliness of our macroeconomic statistics is not yet adequate for anticipating, preventing or moderating the effects of unexpected events such as cyclical downturns.

As already mentioned in chapter 1 and chapter 2, there are three main approaches to be followed for improving timeliness:

- **The speeding up of regular production processes:** Under this approach each stage of the process leading to the production of statistical indicators would need to be redesigned and the whole production process re-engineered. Timeliness is seen as an issue pertaining to the entire production process - hence it has to be addressed by the overall way a survey is organized, there cannot be a simple statistical solution to it. This approach would require large financial investments in the data collection and processing phases and will not realistically produce concrete results in the short/medium term.
- **The incomplete sampling approach:** This approach can consist in identifying incomplete or partial samples for a given survey at disaggregated geographical level. As an example, we can assume to identify incomplete samples for each country of the Euro area or for each state of the US, or even each region of Spain. Those incomplete or partial samples will be not necessarily representative at the level of each country, state or region, but their aggregation should lead to a representative sample for the Euro area, the US or Spain.
- **The model-based approach:** Statistical models can also lead to the production of timely estimates. This would be the cheapest way because it does not call for resources during the data acquisition stage; it would only require an investment related to methodological research and information processing.

The content of this chapter is mainly focusing on statistical and econometric techniques used in producing rapid estimates. Statistical and econometric techniques described in this chapter can be grouped into three main categories.

- Method based on techniques developed within time series analysis, which exclusively use estimates calculated on data without any hypothesised theoretical model suggested by economic theory (Box and Jenkins (1976)). This method uses only the estimates made on previous occasions (if any). Methods that find specific sub-series within a time series, thereby permitting a good prediction of the overall time series to be made (van Garderen et al. (2000); Maravall (1983); Battaglia and Fenga (2003)) can also be placed within this category;
- Methods which use both information obtained from units responding within a fixed time (timely respondents) and the estimates made on previous occasions. These models can be further classified into (i): Methods based on the imputation of the variable values of the non-respondent units. Such methods can be based on a non-parametric approach (Chen and Shao (2000)), or on a suitable model (Little (1986)); a relevant set of models are the linear dynamic models (Harvey and Pierse (1984) and Harvey (1989); Tam (1987); Bell and Hillmer (1990)); (ii) Methods based on re-weighting techniques that correct weights assigned to timely respondents, so that they can also represent non-timely units in a

suitable way. These weights can be based on super population models explaining either the stochastic process that generates the timely response, or the values of the target variable of the non-respondent units at the current time (Rizzo et al. (1996); Eltinge and Yansaneh (1997)); the probability of timely response could also be defined using non-parametric techniques (Giommi (1987); Niyonsenga (1994));

- Suitable econometric methods based on the use of the relationships between the target variables and proxy indicators used as early indices of the target variables (Bodo and Signorini (1987); Bruno and Lupi (2004)). This category includes a large variety of methods from the regression based ones to the multivariate time series models.

The aim of this chapter is to provide a general non-technical review of the models included in the three categories mentioned above. The following chapters of this section will provide deeper and more technical descriptions of some of the models briefly introduced in this chapter, as well as more advanced ones.

5.2 Methods based on univariate time series analysis: brief review and notation

A time series is a variable that is indexed by time. Therefore if we consider x as a variable (univariate case), a time series is denoted by

$$(x_0, x_1, x_2, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_T),$$

considering the time series starting in 0 and ending in T (standard number of observation in a time series). These data are collected over the time as weekly, monthly, quarterly, and yearly values. Usually the researcher would like to understand whether there is some pattern in the values collected to date, with the intention of forecasting. In this section we are describing the various steps in the univariate time series modelling based on the parametric or modern time series approach as proposed by Box and Jenkins (1976). A more extended presentation of those topics can be found in Cochrane (2005) and Hansen (2015).

5.2.1 Data transformation

Stationarity

It is well known that we can define a strict stationary series x_t when its probabilistic characterization remains invariant if it is translated on the time (i.e. if we consider the series x_{t+k}). Computational reasons and the frequent (perhaps abused!) assumption of Gaussian distributions lead to the definition of a weaker stationarity in covariance. A time series is stationary in covariance if its first and second moments are time invariant. No requisites are asked for the other moments and this justifies the weakness of the definition with respect to the strict case[1]. Those series are also called stationary of the second order. It is important to observe that this kind of stationarity is mostly relevant only under the hypothesis of a Gaussian generation process of our series. The explanation for this is that the Gaussian distribution is univocally identified by its first two moments, which is not the case for any non-Gaussian distributions.

Lag Operator

Consider a variable x_t at time t. The lag-operator L (or backward operator B) is defined according to: $Lx_t = x_{t-1}$. Thus, a single application of this operator to a sequence $[x_t]$ define a sequence $[y_t]$ such that $y_t = x_{t-1}$.

Similarly, $L^2 x_t = L(L_x) = L(x_{t-1}) = x_{t-2}$; more generally, $L^s x_t = x_{t-s}$ for $s = 0, \pm 1, \pm 2, \ldots$. When s assumes negative values the operator L^s is called *lead* (or *forward*) operator. The lag (or lead) operator

[1] On the other hand, covariance stationarity does require that all first and second moments exist whereas strict stationarity does not. In this one respect, covariance stationarity is a stronger condition.

is linear. We have $L(K) = K$ if K is constant, $L(ax_t) = aL(x_t) = ax_{t-1}$ for any coefficient a and $L(ax_t - by_t) = ax_{t-1} - by_{t-1}$ as consequence of the linear property.

It is also possible to define a finite or infinite order polynomial according to: $a(L) = a_0 + a_1L + a_2L^2 + \ldots$..The multiplication of a sequence $[x_t]$ by the polynomial $a(L)$ produces the distributed lag sequence $y_t = a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \ldots$.

The use of the lag operator is mainly theoretical, from the practical point of view, lagged variables are inserted in the regressions to model their dynamic. The choice of the power s is one of the most important issues of the regressor selection.

When modelling a time series, differencing and logarithmic, or Box-Cox, transformations are used to make the series covariance stationary. In this case, it is better to transform the series before using a difference operator.

Logarithmic Transformation

The logarithmic transformation is the most widely used. Its effects can be summarized in three directions. For the series x_t we obtain the stabilization of its variance; regarding the regression the log transformation linearizes the increasing slopes of x_t with respect the explanatory variables; finally, the positively skewed distributions of x_t are normalized.

The log transformation is a particular case of a more general class of transformations called *Box-Cox transformation* given by $\frac{X_t^\lambda - 1}{\lambda}$, $0 \leq \lambda \leq 1$; The logarithmic transformation corresponds to $\lambda = 0$. The Box-Cox method permits a more appropriate transformation of the data but, contrary to the logarithmic transformation, it is sometimes difficult to give interpretation of the transformed data on the basis of the estimated parameter λ. (more details on transformation of variables can be found in (Johnston (1984)) pp. 61-74).

Difference Operator

Consider a variable x_t at time t. The difference operator Δ is defined according to: $\Delta x_t = x_t - x_{t-1}$. We have $\Delta = 1 - L$ and in general $\Delta^d = (1 - L)^d$. For example, $\Delta^2 x_t = x_t + x_{t-2} - 2x_{t-1}$. The purpose of the difference operator is to make the series stationary in mean. Attention must be paid to the excessive use of the difference transformation. In a properly specified regression equation overdifferencing the series induces a non-invertible moving average (MA) process in the disturbance of the transformed regression (Plosser and Schwert (1977)) and can increase the variance of the stationary series. For some non-stationary series it is not suitable to proceed directly with the difference operator and it might be preferable to proceed first with a transformation of the data, say, a logarithmic transformation. Suppose that the series x_t has an exponential trend, then no difference degree d will set it to a constant, since $\Delta^d(e^{\alpha t}) = (1 - e^{-\alpha})^d e^{\alpha t}$. In this example to obtain a transformed stationary series, the appropriate operation is $\Delta log x_t$.If the series follows an exponential process; this means that it is characterised not only by non-stationarity in mean but also in variance. The variance of the series will grow with the time. The logarithmic transformation will reduce or cancel the non-stationarity in variance so that the Δ log series will appear to be stationary both in mean and in variance (second order stationarity).

Seasonal difference

Other operators as $\Delta_s = 1 - L^s$ with suitable integer value of s can be useful to filter the seasonal component of the series. The most used operators are Δ_{12}, Δ_4 for the monthly and quarterly data respectively. As before, if a series is characterized by a non-stationarity in variance, just the seasonal difference will be not enough to achieve the stationarity so that it will need to be combined again with a logarithmic transformation.

5.2.2 Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The Autocorrelation Function and Partial Autocorrelation Function summarize several aspects of time series and are strictly linked with the concepts of stationarity, autocovariance and partial autocovariance functions and the different classes of time series.

Autocovariance Function

A covariance stationary time series x_t has, by definition, the following properties:

$$\begin{aligned} Ex_t &= \mu \\ \gamma_k &= E(x_t - \mu)(x_{t+k} - \mu) \end{aligned} \tag{5.1}$$

where k is the lead (lag) shift of the series if k is positive (negative). γ_k is the autocovariance function and, as the mean value μ, does not depend on time but only on the shift k. Both μ and γ_k take any value in the real axis when $k \neq 0$. If $k = 0$, then:

$$\gamma_0 = E(x_t - \mu)^2 = \sigma^2 \tag{5.2}$$

represents the variance of the time series and assumes only positive values. The autocovariance function gives information on the variability of the series and on its intrinsic dynamics and consequently on the dependency of present observations from past ones. A series with zero covariance from all leads and lags different from zero, is called white noise and it is characterized by no links between present and past observations so that it is unforecastable. If we wish to compare the time link behavior within two or more series then we need the concept of correlation.

Autocorrelation Function and Partial Autocorrelation Function

The autocorrelation function (ACF) of a stationary time series x_t is defined as:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{5.3}$$

The ACF takes real values between -1 and $+1$, with value near ± 1 indicating a stronger correlation. The ACF loses the information on the variability of the time series.

Another way to measure the correlation between the process at different points in time (for instance between x_t and x_{t+k}) is to make a regression of x_t on a constant and the sequence $x_t, x_{t+1}, \ldots, x_{t+k-1}, x_{t+k}$, and take as partial autocorrelation the regression coefficient related to x_{t+k}; It calculates the correlation of x_t with x_{t+k} removing all the correlation for the intermediate lags.

In other words, filter out x_t and x_{t+k} the linear influence of the variables of the process that lie in-between, $x_{t+k}, \ldots, x_{t+k-1}$, and calculate the correlation of the transformed variables. This is called Partial AutoCorrelation Function (PACF).

Classes of time series

If a time series x_t is covariance stationary then the Wold decomposition theorem applies[2]:

$$x_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t-j} \tag{5.4}$$

[2]Here for sake of simplicity we assume a time series x_t with zero mean.

where ϵ_t is a white noise process[3]. The interpretation of equation 5.4 is straightforward: a stationary time series has an exact approximation by a linear stochastic process. In other words, the class of linear processes contains all the covariance stationary processes, even the highly non-linear.

In the same class we have particular finite lag processes as the Moving Average processes MA(q) defined as:

$$x_t = \sum_{j=0}^{q} \beta_j \epsilon_{t-j} \tag{5.5}$$

the Autoregressive processes, AR(p):

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \epsilon_t \tag{5.6}$$

and the mixed Autoregressive Moving Average processes ARMA(p,q):

$$x_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \sum_{j=0}^{p} \epsilon_{t-j} \tag{5.7}$$

The ACF is informative for the order q of the MA processes and is not for the order p for the AR processes. On the contrary, the PACF is informative for the order p of the AR processes and is not for order q of the MA processes. ACF and PACF are not informative for the orders p and q of the ARMA processes and in this case other statistics are needed (For example, see Gourieroux and Monfort (1997), pp. 164-167).

Nevertheless, we can consider that both ACF and PACF are essential tools in the identification of the univariate time series models as suggested by Box and Jenkins (1976).

5.2.3 Periodogram: useful for frequency or cyclical analysis

The periodogram is a useful tool for the estimation of the spectral density function. The *spectrum* or *spectral density function* of a time series process is the real-valued function defined as follows:

$$h_x(w) = \frac{1}{2\pi}\left(\sum_{k=-\infty}^{\infty} \gamma_k e^{-iwk}\right), \forall \omega \in \mathbf{R} \tag{5.8}$$

where γ_k is the autocovariance function, $\sum_{k=-\infty}^{\infty} \gamma_k e^{-iwk}$ is the evaluation of the autocovariance generating function $g_x(z) = \sum_{k=-\infty}^{\infty} \gamma_k z^k$ at the complex value $z = exp(iw)$ with $i = \sqrt{-1}$ and w a real number. We can compute the spectrum of a time-series process for any value of w if a set of autocovariances γ_k is available.

The spectrum expression could be as follows[4]:

$$h_x(w) = \frac{1}{2\pi}(\gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k cos(wk)), w \in [0, \pi] \tag{5.9}$$

If we consider a sample that has T observations, we could compute the autocovariance function and the variance for that sample[5].

[3] A white noise process is characterized by zero mean, constant variance and zero autocovariance for all shifts k except for $k = 0$.

[4] Such a result can be obtained by assuming: symmetry of autocovariance function, symmetry of $\sin$ and $\cos$ functions, by applying the De Moivre's Theorem and the property of polar values.

[5] The sample autocovariance function is $\hat{\gamma}_k = \hat{\gamma}_{-k} = \frac{1}{T}\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})$ and $\bar{x} = \sum_{t=1}^{T} x_t$ for $k = 0, 1, \ldots, T-1$

According to a particular sample, the periodogram is computed as follows:

$$\hat{h}_x(w) = \frac{1}{2\pi}(\hat{\gamma}_0 + 2\sum_{k=1}^{T-1}\hat{\gamma}_k cos(wk)) \tag{5.10}$$

where $\hat{\gamma}_k$ is *the sample autocovariance function*. The periodogram is an estimator of the spectrum and it is a useful tool for frequency and cyclical analysis.

Consider a time series with $M = \frac{T-1}{2}$ observations and frequencies $w_j = \frac{2\pi j}{T}$ for $j = 1, 2, \ldots, M$; The lower frequency observed might appear once in the whole sample period $w_1 = \frac{2\pi}{T}$, on the other hand the highest would be $w_T = 2\pi$. Taking integrals of the w_j periodogram functions, the variance of sample could be computed. In fact $2\int_0^{w_j}\hat{h}_x(w)dw$ defines the sample variance for cycles within w_j frequency.

In practice, if the periodogram density is high for low frequencies, most part of the sample variance is concentrated within low frequency cycles and the long run dependencies are very strong. It can be shown that the autocovariance function and the spectral density are equivalent, both are non-parametric description of the main features of the series x_t, the former in the time domain and the latter in the frequency domain. From the theoretical point of view, spectral analysis is preferable in order to simplify the achievement of some results[6]. In applied analysis, the choice to use one or the other is a matter of convenience in the representation of the behavior of the series and upon the fact that the spectral density is very sensitive in revealing its cyclical components. Unfortunately, the periodogram is not a consistent estimator of the spectrum. Consistent estimators, called spectral windows, have been proposed in the literature usually based on smoothed versions of the periodogram (Bartlett (1948); Parzen (1957);Tukey and Blackman (1959); etc.).

5.2.4 Plot of a single or set of aggregate or indicator series

Graphical analysis is a key milestone when you want to model a series or to build a rapid estimate model involving several series. This exploratory data analysis permits discovering structures in a time series (presence of trend-cycle, outliers, seasonality, trading-day effect) or studying the links between several series.

Exploratory data analysis of a time series

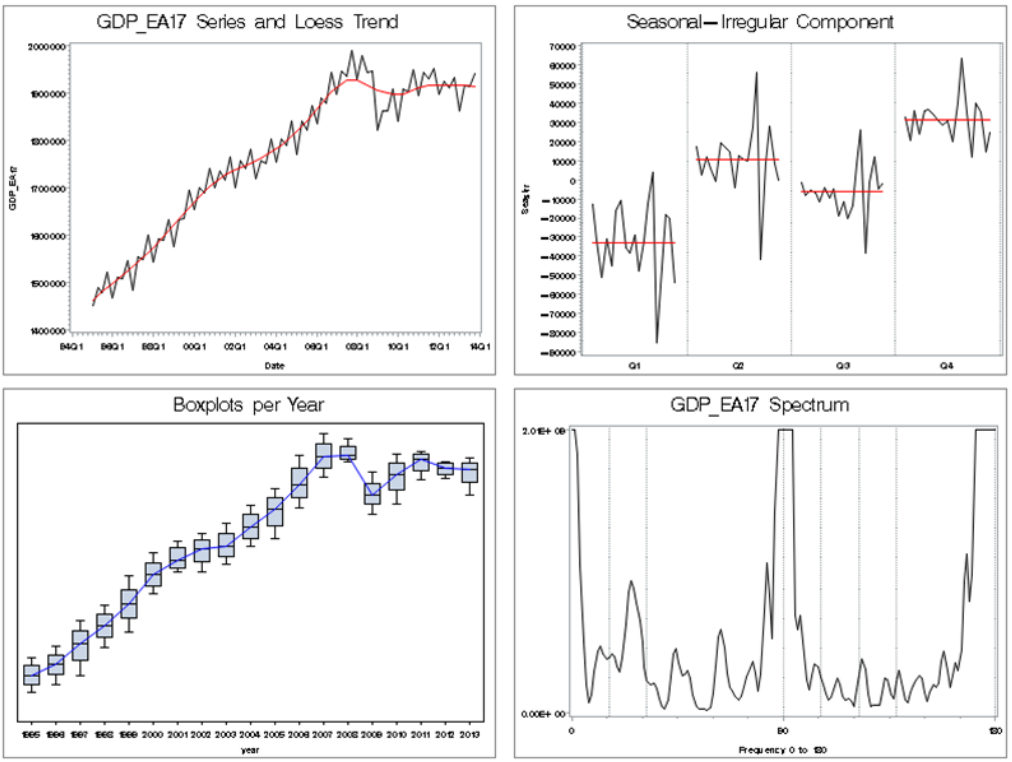
For a series, this exploration can be done using simple graphs based on rough estimations of the components. A basic set of graphs could be:

- A graph of the raw data with an estimation of the trend-cycle component. This very simple graph will show the presence of a trend-cycle, of possible outliers and of seasonality.
- A graph of the seasonal-irregular component, obtained as the difference between the raw series and the trend-cycle estimate, by month (or quarter). This graph should give you a better idea of the presence and characteristics of a seasonal component.
- A graph of the boxplots by year. The graph usually gives an idea of the stationarity in mean and variance, and shows local outliers.
- A graph of the series spectra which usually shows the presence of a trend- cycle (low frequencies), seasonality and trading-day effect.
- And graphs of some ACF to help you in the modeling of the series: ACF of the raw data, ACF of the first difference, ACF of the seasonal difference and ACF of the first and seasonal differences.

Figure 5.1 presents some of these graphs for the Euro area GDP series. These simple graph reveal a lot of useful information:

[6] For example, the convolution integrals in the time domain can be substituted by the simplest products in the frequency domain.

Figure 5.1: Graphical exploration of a time series (Euro area GDP)



- The usual graph of the series shows a strong trend: the series is not stationary in mean;
- The boxplots show no evident evolution of the series variance across time and no additive outliers;
- The series is clearly seasonal as shown by the spectral peak at frequency $\frac{\pi}{2}$. This seasonality is characterized by a strong fourth quarter and a weak first quarter;
- A clear rupture appears in 2009 that will have to be taken into account if one wants to fit an ARIMA model to the series.

Looking for relationships between several series

When the number of variables is quite small, let us say up to a dozen, very simple graphs can really help to reveal relationships: overlapped graphs of the series, sometimes after a suitable transformation like standardization, or representation of the cross-correlation functions.

When series are not stationary, usual graphs of raw data are often misleading. It is sometimes better to work on stationarized series, after proper differencing and log-transformation. Comparing growth rates is often a good idea.

Figure 5.2 compares the quarterly IPI and GDP of the Euro area. If the two series tell us roughly the same story, their correlation is definitely clearer on the growth rates.

As confirmed by the cross-correlation function (Figure 5.3), the two series are coincident: their correlation is maximal at lag 0. Anyway, as the IPI is a monthly indicator, it is available before the GDP. The first estimate of the Euro area GDP is available 45 days after the end of the quarter when the IPI is available 40 days after the end of the month. It means that for a given quarter, the IPI of the 2 first months are available 10 days after the end of the quarter which gives a quite good idea of the quarter under review. This very simple graphical analysis shows that the IPI is a good candidate for a rapid estimation of the Euro area GDP.

Figure 5.2: Comparison of the Euro area GDP and IPI

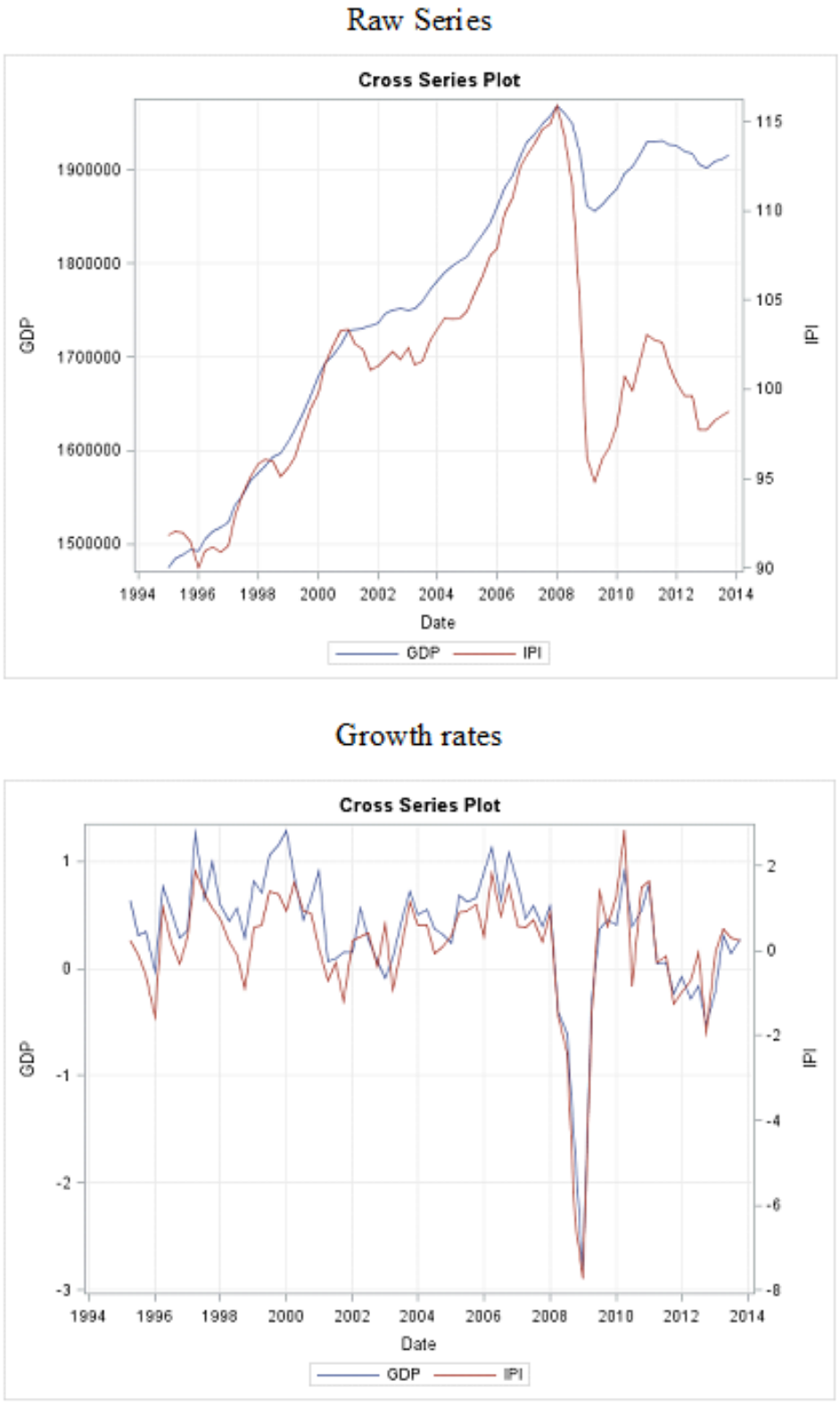
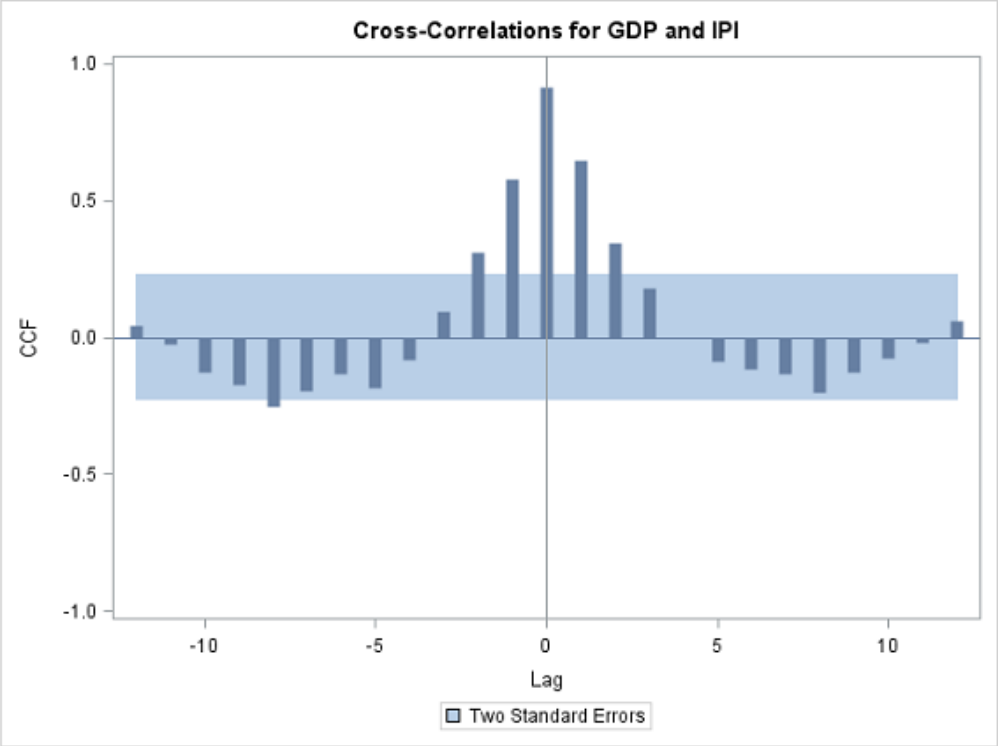


Figure 5.3: Cross-correlations between the Euro area and GDP



When the number of variables to compare is large or even very large, other graphical techniques must be used. Among them, cluster analysis and principal component analysis, that will point out series sharing a similar dynamics, can be mentioned.

5.2.5 Trend stationary (TS) vs. Difference Stationarity (DS)

We consider an economic time series variable x_t. This variable may be composed of unobservable components such as: *deterministic trend* (d_t), *stochastic trend* (s_t) and *stationary component* (c_t).

We can distinguish between *trend stationary* (TS) and *difference stationary* (DS) hypotheses on the basis of the elements that compose the economic series taken into consideration.

For the TS hypothesis the composition of x_t is: $x_t = d_t + c_t$, for the DS hypothesis the composition also includes the component s_t. The presence of the stochastic trend creates substantial differences between TS and DS.

First, the deterministic trend (if any) cannot be interpreted as a central curve towards which the actual values of the series revert after some temporary deviations from it. Second, the application of standard tests, e.g., Student t, can substantially increase the error probability, therefore they are no longer applicable. Finally, the consequences for the forecasts are different. On the one hand, if we suppose that series x_t is integrated of order one, i.e. I(1), then the forecast-error variances grow linearly in the forecast horizon, because the presence of the unit root indefinitely accumulates previous disturbances. On the other hand, if x_t is TS, then the forecast-error variances are bounded, because the conditional h-step ahead forecast error, given the deterministic component, is simply the disturbance term for that period.

5.2.6 Seasonally adjusted vs. seasonally unadjusted data

Some researchers have taken into consideration the possibility that particular events occurring seasonally or in given periods of time could influence the estimation and forecasting power of a model. For instance, Bovi et al. (2000) adopted the *Tramo-seats* procedure and Camba Mendez et al. (2001) purge the data of any outlier that does not have a clear economic interpretation. Others have used data without any outliers elimination, such as Statistics Sweden (2000). In that case they have preferred the seasonally unadjusted data approach. In constructing rapid estimates, the choice between seasonal and non-seasonal adjusted data is not only a problem of personal preferences but it is strongly influenced and guided by the characteristic of the target variables and by its relevant transformation for policy making purposes.

For example, concerning inflation, the annual inflation rate is the most relevant transformation to which users are looking at. In this case, the rapid estimates should aim to provide an early and reliable estimate of this variable. Since the annual inflation is measured as a log difference over 12 months of the non-seasonally adjusted price index, when building up rapid estimates models, we should use non-seasonally adjusted price data even when seasonally adjusted ones are available. By contrast, since the most relevant transformation of the GDP is its quarterly growth rate measured from seasonal and, possibly, calendar adjusted data, the rapid estimates models should be based on them and not on unadjusted data or simply calendar adjusted data, even when they are equally available.

5.2.7 Most relevant univariate time series models

The modern time-series analysis considers a given individual observed time series as a realization of an unknown stochastic process generating it. Since socio-economic phenomena cannot be replicated, the identification of the underlying generating process of an observed time series appears almost impossible from a unique realization, unless very restricted identification constraints are imposed. To overcome this situation, it is usually assumed that the process is "ergodic" so that its moments, like mean and variance, can be derived from a single realization of the process. This because, under ergodicity conditions, the moments estimated from a single realization are equal to the ones computed on each probability distribution characterizing the process at each point of the time scale over which it is defined. This condition is only applicable to Gaussian processes, meaning these based on normal distributions with finite mean and variance. Under these two hypotheses, the autocorrelation and partial autocorrelation function can provide useful information about the structure of the typology of the models which fit the data. This approach was formalized for the first time by Box and Jenkins (1976) and it is nowadays considered the basis for time series modeling. Below we are shortly introducing some of the most relevant models commonly used in theoretical and empirical works, with particular attention to those having special relevance in nowcasting exercises.

Moving average models

In 1937 Slutsky demonstrated that a weighted linear combination of purely random variables produce an autocorrelated series, which has some similar characteristics with most economic indicators, opening the door to the definition of the so called moving average models. We define such a model as: $Y_t = \theta(L)\epsilon_t$ where $\theta(L)$ is an infinite polynomial in the lag operator L and ϵ_t is a white noise with 0 mean and finite variance σ^2. Since this model is defined over an infinite horizon, its relevance is mostly theoretical but it cannot have any empirical applications. By contrast, it is possible to consider a finite order MA (q) moving average model by truncating the infinite order after, let's say, q lags: $y_t = \theta_q(L)\epsilon_t$, where $\theta_q(L) = (1 + \theta_1 L + \theta_2 L^2 + \ldots \theta_q L^q)$ and ϵ_t is as defined above.

Autoregressive models

Let us consider the finite order moving average MA(q) and assume that all characteristic roots of the polynomial equation $\theta_q(L) = 0$ lay out of the unit circle. This particular situation is usually referred as invertibility

condition. If this is the case so that the moving average is invertible, we can write the previous equation as an infinite autoregressive form:

$\phi(L)y_t = \epsilon_t,$

where $\phi(L)$ is an infinite order polynomial in the lag operator L. Once again this model has not empirical relevance but this is not the case of its truncated version AR (p) defined as:

$\phi_p(L)y_t = \epsilon_t,$

where $\phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots \phi_p L^p)$. If the characteristic roots of the polynomial equation $\phi_p(L) = 0$ lay out of the unit circle, the model is called stationary and it can be transformed into an infinite moving average one.

ARMA models

Box and Jenkins (1976) proposed to combine the autoregressive part with the moving average one into a unique model in order to have a class of models able to capture in a better way all features of real variables. Furthermore, the ARMA models could be, in most cases, more parsimonious than of corresponding AR and MA. An ARMA model is usually referred to as ARMA (p,q) where p is the order of the autoregressive part and q of the moving average one.

$\phi_p(L)y_t = \theta_q(L)\epsilon_t$

or:

$y_t = (\theta_q(L)/\phi_p(L))\epsilon_t$

Non-stationary models (ARIMA)

All models presented until now assume that the variable y_t is stationary in mean while we have assumed that a logarithmic transformation has already been applied to eliminate heteroschedastic effects. In case where Y_t is characterised by a stochastic trend, it has to be subject to a differentiation process to achieve the stationarity before fitting the autoregressive, moving average and ARMA models. By combining the previous specifications, we obtain the following integrated versions:

- Model ARI (p, d): $\phi_p(L)(1 - L)^d y_t = \epsilon_t,$

  where p is the order of the autoregressive part and d is the differentiation order.

- Model IMA (d,q): $(1 - L)^d y_t = \theta q(L)\epsilon_t,$

  where d is the differentiation order and q is the order of the moving average part.

- Model ARIMA (p, d, q): $\phi_p(L)(1 - L)^d y_t = \theta_q(L)\epsilon_t$, or: $(1 - L)^d y_t = (\theta_q(L)/\phi_p(L))\epsilon_t,$

  where d is the differentiation order and p and q the autoregressive and the moving average order, respectively.

Random walk models

This is a very simple and even trivial class of models, without any typical autoregressive or moving average component. The basic random walk model is expressed as: $Y_t = Y_{t-1} + \epsilon_t$ where ϵ_t is an uncorrelated zero mean random variable with finite σ^2 variance.

The interpretation of the model is that the today's expected level of a given variable is the same as yesterday plus a random disturbance. The main variant of this equation is the integrated random walk which applies to the case where ΔY_t is not stationary in mean. In this case, the random walk will become: $\Delta Y_t = \Delta Y_{t-1} + \epsilon_t$

With this formulation, we hypothesize that the expected today's growth of a given variable is the same as yesterday plus an error disturbance. Indeed, it is not unlike to assume, under normal economic conditions, that this month's inflation will be the same as the previous month plus an unknown disturbance.

SARIMA models

When in an ACF seasonal lags at lag 12, 24, 36 etc. (or 4, 8, 12 etc.) are statistically significant, the series can be considered as seasonal. Instead of modeling all the ACF lags together, we can factorize the model splitting the seasonal and non-seasonal parts.

SARIMA$(p,d,q)(P,D,Q)_S$ models are used for non-stationary time series, with trend and seasonality at lag (S), and are defined by the following general model:

$$\phi_p(L)\phi_P(L^S)(1-L)^d(1-L^S)^D Y_t = \theta_q(L)\theta_Q(L^S)\epsilon_t$$

One of the most popular model, which can fit a large part of economic series, is the "Airline model", a very parsimonious model with only 2 parameters, that corresponds (in case of monthly data) to the SARIMA model $(0,1,1)(0,1,1)_{12}$:

$$\Delta\Delta_{12}Y_t = (1+\theta_1 L)(1+\theta_2 L^{12})\epsilon_t$$

Univariate models and rapid estimates

It is quite obvious that univariate models should not be used to produce any kind of rapid estimates, unless the extreme situation where no additional related indicators to the target variables can be found. Alternatively, univariate models can appear to be the best solution if the related indicators found are so unreliable or of so poor quality to compromise the forecasting exercise. What said before could convince the reader that univariate models can never produce good forecast and that a multivariate model with reliable information could provide best forecast in most circumstances. Empirical studies show that this is not really the case because it is often hard to beat simple univariate models by multivariate ones within a one step ahead forecasting exercise. This is particularly true during periods characterized by a regular evolution of economic activity, while in presence of turning points, structural breaks and other significant changes, the limitations of univariate models will appear in a clear way. In particular, the most commonly used univariate models during a one step ahead forecasting exercise are the simple AR (1) and the random walk, also in their integrated versions. Due to their good properties, they are usually considered as the ideal benchmark when developing more sophisticated forecasting models.

5.3 Rapid Estimates based on Factor Models

Factor analysis is a statistical method used to summarize the variability among a set of observed, correlated statistical variables in terms of a potentially lower number of latent variables called factors. The main assumption underlying the factor analysis is that there are some common driving forces which characterize a set of variables which are expressed by the factors. Factor analysis has been used for years as a powerful instrument for data reduction. More recently, thanks to the pioneer work of Stock and Watson (1989), it has also been used like a powerful instrument for nowcasting and short-term forecasting. This section is introducing the basic concepts of static and dynamic factor analysis which are treated in detail in chapters 10 and 11. For further details on this topic, please refer to Stock and Watson (2010a) and Barigozzi et al. (2014).

5.3.1 Data Selection

In order to explain past economic growth and, on the basis of that, evaluate a good forecasting strategy, researchers must not only find an equation which happens to fit the data, but they have to find the best forecasting equation. Sometimes they do not take into consideration this postulate and some indicators have

been chosen because they explain a significant amount of the GDP variation about its mean. Bovi et al. (2000) set down some criteria that a regression should satisfy to be considered for forecasting purposes. For instance, residual autocorrelation test, ARCH test, heteroscedasticity test, normality test and parameter stability (Chow) test for an out of sample period. In addition, an alternative approach is offered by Blake et al. (2000).

5.3.2 The Blake, Kapetanios and Weale two-stage approach

Let there be m available indicators for forecasting a given series. We consider this simple regression:

$y_t = \sum \beta_i x_{it} + \epsilon_{it}, t = 1, \ldots, T - \tau,$

where x_{it} are the indicators, $t = 1, \ldots, T$ and $i = 1, \ldots, m$ and the period starting at $T - \tau$ and ending in T denotes the forecasting period. We choose the residual sum of squares (RSS_i) in order to measure the fitting powerfulness of x_{it} on y_t. We rank the indicators in ascending order of RSS_i, therefore the indicators are ranked in terms of best fit. Then the q best-fitting indicators are chosen for a further round of model building.

All possible combinations of the q indicators are used as possible models, so it should be a good solution that q will not be very large. The set of all possible combinations of the q indicators is Γ .

Then the combination j ($j \in \Gamma$) that minimises the Akaike information criterion is chosen to produce the forecasts over the forecast evaluation period.

The reason for considering a two stage approach for forecasting model building is that considering all possible combinations of all possible indicators is computationally prohibitive.

5.3.3 Data reduction: factor analysis

In the context of factor analysis we assume that, given a large dataset describing the available information, there are a small number of combinations of original variables (factors) which describe the common behaviour of the dataset and explain a large amount of the variability of the dataset itself. Factor analysis and similar approaches are usually referred as data reduction techniques. They reduce the variable space but they do not select variables, since identified factors or components are linear combinations of the original variables.

Static factor analysis: principal components

Let us consider the m indicators

$$X = \begin{bmatrix} x_{t1} & \ldots & x_{tm} \end{bmatrix} = \begin{pmatrix} x_{11} & \ldots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{T1} & \ldots & x_{Tm} \end{pmatrix}$$

The indicators have to be measured in the same units. The main aim of the principal component analysis consists in extracting from m indicators a subset of variables that account for the most, or all, of the indicators' variability. The principal components are a linear combination of the m indicators where the first principal components accounts for most of the variation in the m indicators, the second principal component accounts for most of the remaining variability in the m indicators and so on.

We define the first principal component as the T-vector z_1 which components are: $a_{11}x_{t1} + a_{12}x_{t2} + \cdots + a_{1m}x_{tm}$, with $t = 1, \ldots, T$, or $z_1 = Xa_1$, where a_1 is a m-vector of parameters. We wish to choose a_1 so that z_1 accounts for the most of the variability in the m indicators. Hence, we choose a_1 to maximize the sum of squares of z_1, $z_1'z_1 = a_1'X'Xa_1$

Taking into consideration the normalization $a_1'a_1 = 1$, the problem is reduced to the maximization under the normalization constrain[7].

Subsequently, we define $z_2 = Xa_2$ in order to choose a_2 to maximize $z_2'z_2 = a_2'X'Xa_2$ subject to two constraints: normalization ($a_2'a_2 = 1$) and another constraint that ensures uncorrelation between z_1 and $z_2(a_1'a_0)$[8].

The Lagrange multipliers associated to the normalization constraints (so called λ_1 and λ_2) must be chosen as the first and second largest eigenvalues of $X'X$[9]. We proceed in this way for each of m eigenvalues of $X'X$; The resultant eigenvectors are assembled in an orthogonal matrix:$\begin{bmatrix} a_1 & a_2 & \dots & a_m \end{bmatrix}$

The m principal component of X are given by matrix $Z = XA$ ($T \times m$ dimensions). Furthermore,

$$Z'Z = A'X'XA = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix}$$

shows that the principal components are uncorrelated and $z_i'z_i = \lambda_i$ for $i = 1, \dots, m$.

We define m as the total variation of the indicators under analysis and λ_i as the variation accounted for by the specific i^{th} principal component.

Since the total variation m is given by the sum of the singular variations of specific principal components ($\lambda_1 + \lambda_2 + \dots + \lambda_m = m$), $\frac{\lambda_i}{m}$ represents the proportionate contribution of the i^{th} principal component to the total variation in the m indicator variables.

Dynamic factor analysis

Let us consider the state-space form Engle and Watson (1981):

$$f_t = \phi f_{t-1} + \gamma z_t + v_t \tag{5.11}$$

$$x_t = \alpha f_t + \beta z_t + e_t \tag{5.12}$$

where $\begin{bmatrix} v_t \\ e_t \end{bmatrix} \sim N.I.D \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}$ and x_t is the m vector of indicator variables, f_t is a k vector of unobservable factors and z_t is a vector of exogenous variables. The most popular models were popularized by Stock and Watson (1989).

They extracted from the set of indicator variables a latent variable that could be identified as the state of the economy; they then forecasted this variable using a quite separate vector autoregressive (VAR) model, focusing on the chance of a recession. A comprehensive overview of the utilization of dynamic factor models is provided in chapter 10.

Forecasting by means of the Stock and Watson factor model

Stock and Watson (1998) and Stock and Watson (2002) postulated that aggregates and indicators have some dynamic factors in common. They came up with a model where the one step-ahead prediction of aggregate

[7] The maximization problem is set up by using Lagrange multipliers formula associating (λ Lagrange multipliers) to the normalization constraint.

[8] Another maximization problem is set by using two different Lagrange multipliers (λ for the new normalization constraint and μ for the uncorrelation constraint between z_1 and z_2)

[9] For instance, the optimal solution for the first maximization problem is $(X'X)a_1 = \lambda_1 a_1$that conduces to $z_1'z_1 = \lambda_1 a_1'a_1 = \lambda_1$

and dynamic indicators are treated. Prediction of aggregate depends upon the current and past value of it, idiosyncratic error and dynamic factors. Indicators depend mostly on dynamic factors and, of course, on an idiosyncratic error. Therefore, dynamic factors play an important role both in indicators and aggregate specification. Stock and Watson provided some techniques for the estimation of these factors. For the purpose of this handbook, we are obviously focusing on nowcasting instead of forecasting the target variable[10]. Firstly, estimation of dynamic factors from indicators should be done. After doing so, estimation of parameters of nowcasting equation have to be performed. Finally, the aggregate nowcast would be provided. Nevertheless, there is no *a priori* reason why the best explanatory factors of overall variation should be also the best in forecasting the particular value of aggregate. It may be more interesting to take into consideration the factors that best explain aggregate. For example, we could choose the smallest set of factors which explain a fixed percentage of the overall aggregate variation[11]. A detailed description of the Stock and Watson approach is presented by Frale et al. (2011).

The Camba-Mendez, Kapetanios, Smith and Weale Automatic Leading Indicator Model

Camba Mendez et al. (2001) offer an extension of Stock and Watson's work, the so called Automatic Leading Indicator Model.

In this model instead of using all the observable exogenous variables x_t, it is assumed that their influence on the target variable y_t (variables of interest) may be conveniently summarised via k driving forces or *factors* s_t which underlie the fluctuations of the variables x_t .

In order to carry out this procedure we need a two-stage process.

- In the first stage, a Kalman filter technique can be used to estimate the value of s_t from observations on x_t[12].
- In the second stage, the factors s_t could be incorporated into a VAR model to forecast y_t via $a_y(L)y_t = a_{S(L)}s_t + e_t$ [13]
- Given the $[s_t]_{t=1}^{T}$, the unknown coefficients $[a_y]_{i=1}^{P_y}$ and $[a_s]_{i=1}^{P_s}$ may be estimated by least squares over the sample period, $t = 1, \ldots, T$.

5.4 Rapid estimates based on regression models

Regression models describe the relationship between one or more variables (also called dependent variables, regressands, explained variables) and other variables (independent variables, regressors, explanatory variables). Such models have been originally defined in the linear context but they can also be generalised to the non-linear one. In this section, we are mainly focusing on linear specifications even if we are allowing for non-linear data transformations such as log and Δ log ones. We are providing here the basic concepts of static and dynamic regression models which have been largely used in nowcasting and forecasting exercises. For an in-depth analysis of the linear regression models we refer to Stock and Watson (2010b) and Andren (2007).

[10] Since current value is the central point in nowcasting problems, only past aggregate value are considered.

[11] Factors are chosen on the basis of their R_2 with aggregate.

[12] The dynamic factor model from which we could extract s_t is:

$$x_t = Bs_t + u_t$$

$$C(L)s_t = \eta_t,$$

where B is a $(k \times m)$ matrix of unknown parameters and u_t a m-vector of disturbances with $k \leq m$. By Kalman filter we can extract s_t from observations 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.

[13] $a_y(z) = 1 - \sum_{i=1}^{P_y} a_{yi}z^i$, $a_s(L) = 1 - \sum_{i=1}^{P_y} a_{Si}z^i$ and $[e_t]_{t=1}^{\infty}$ a zero mean, conditionally homoskedastic and serially uncorrelated error process with positive definite variance matrix.

5.4.1 Static and dynamic models

In a static regression model, we assume that a given variable, called dependent variable y_t, is explained by a set of variables, called explanatory variables, or regressors, $X_t : y_t = f(X_t, \beta) + u_t$, where β is a vector of unknown parameters to be estimated and u_t is a vector of errors summarizing the part of the variance of y_t not explained by the set of regressors X_t. In this case, we assume that the relationships between X_t and y_t is purely static and its effect operates only within the period. By using static models, a researcher can study how variables affect each other at the same time. For instance, we are concentrating our attention on the simple linearized version of the equation above: $y_t = \beta_0 + \beta_1 x_t + u_t$, where the dependent y and explanatory variable X and the error term u are considered at the same period of time. We assume that u is normally and independently distributed with zero mean and a finite and constant variance $\sigma^2 : u \sim NID(0, \sigma^2)$. The regression parameters β_0 and β_1 are respectively the intercept and the slope of the model and, under the assumption that the regressor x_t is deterministic, they can be estimated equivalently either by maximum likelihood (ML) or by ordinary least squares (OLS), which are the best linear unbiased estimators (BLUE). We can generalize this model to the case of multiple regressors by using the following matrix formulation: $y = X\beta + u$, where y is a $T \times 1$ vector of observations of the dependent variable, X is a $T \times k$ matrix of k deterministic regressors (including a column of one allowing for the presence of the intercept), β is a vector $k \times 1$ of unknown parameters and u is a $t \times 1$ vector of errors so that $E(u) = 0, E(uu') = \sigma^2 I$, where I is the identity matrix. Under such hypothesis, the BLUE estimator of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'y$$

By relaxing the hypothesis of deterministic regression, the OLS are not anymore efficient but they remain unbiased and consistent. By contrast, if we relax the hypothesis of diagonality of the variance/covariance matrix of errors by assuming $E(uu') = \sigma^2\Omega$, where Ω is symmetric and positive definite, the OLS estimator of β is not anymore consistent. In presence of a known diagonal variance/covariance matrix of errors, the unknown parameters of the regression model must be estimated by using the generalized least squares (GLS): $\beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$.

On the other hand, a dynamic model could appear with lagged explanatory variables and/or lagged error term. For example, we can have a model in this general form $y_t = f(x_t, y_{t-1}, y_{t-2}, \ldots, y_{t-p}, u_t)$. In that case, the dependent variables y are related by a particular function to a contemporary independent variable x, a lagged dependent variable and a contemporary error term. The model above can generate very relevant models for the construction of rapid estimates which can assume the following forms:

$$\begin{array}{lcl} y_t & = & \alpha y_{t-1} + \beta x_t + u_t \\ y_t & = & \beta(L)x_t + u_t \\ \alpha(L)y_t & = & \beta x_t + u_t \\ \alpha(L)y_t & = & \beta(L)x_t + u_t \end{array} \tag{5.13}$$

where α and β are unknown parameters, (L) the lag operator, $\alpha(L)$ and $\beta(L)$ are polynomials of unknown parameters and u_t is the error term assumed to be normally and independently distributed with zero mean and finite and constant variance σ^2. Under specific conditions of stationarity and invertibility, the four equations above can be transformed one into the other. In particular, under the stationarity assumption of the polynomial $\alpha(L)$, the third equation can be transformed into the second one with an infinite lag order of x_t. On the other hand, under the invertibility condition of the polynomial $\beta(L)$, the second equation can be transformed into the third one with an infinite lag structure of y_t.

In the dynamic models, users should not consider only the relationship between variables in the same period but also those over the time.

The use of static or dynamic models is matter of specification searches. Economic theory may suggest the choice of variables to relate each other, but rarely suggests the kind of dynamic characterising the models. The identification of the dynamic structure of a model is mainly a data driven process than one suggested

by economic theory. Cross- correlation functions, cross-spectrum, Granger causality, etc., can give insight on the underlined dynamic structure of a given economic relationship. A typology of dynamic models is widely discussed in Hendry (1995)[14] and also in Harvey (1981). The use of a static regression merits some comments. Except in the case when the data generating process is indeed based on a static regression, the estimated residuals cannot be represented by an innovation process and will be generally autocorrelated. Also the forecasts are very poor because the model does not contain the past value of y_t. Finally, a *spurious* regression problem can arise when both y_t and x_t are integrated processes.

The estimators of the parameters are non-consistent in the case in which relevant lagged variables are omitted. Consequently, the interpretation of the $\hat{\beta}_1$ values are misleading. However, the use of static model is relevant in the context of cointegration analysis. If both y_t and x_t are cointegrated process, then $\hat{\beta}_1$ is the estimation of the long-run component and the static model is used as the first of two-step estimation procedure suggested by Engle and Granger (1987).

Both static and dynamic regression models have proven their usefulness in building rapid estimates especially in nowcasting exercises. In particular, the first dynamic model presented above with one or more independent non-lagged variables, has proven to be quite performant and parsimonious at the same time, especially for estimating macroeconomic indicators such as GDP or inflation which are characterised by quite strong entropy.

5.4.2 Long-run component

In the line of the Beveridge and Nelson (1981) decomposition, which provides measures of trend and cycle for an integrated time series, Hatanaka (1996) explains how to define a measure of the long-run component.

Let us consider a stochastic process $\{y_t\}$ that could be stationary I(0) or non-stationary I(1).

We define Δy_t as follows:

$$\begin{aligned} \Delta y_t &= b_0\epsilon_t + b_1\epsilon_{t-1} + \dots \\ &= \sum_{j=0}^{\infty} b_j\epsilon_{t-j} \\ &= b(L)\epsilon_t \qquad \text{with } b_0 = 1 \end{aligned} \tag{5.14}$$

where ϵ_t is an independent and identically distributed (i.i.d.) process with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma_\epsilon^2$; The present time is t and we assume also that the process starts at $t = -\infty$.

Hatanaka shows that the optimal predictor of y_t in the infinitely remote future is:

$$\hat{y}_t = E(y_\infty | y_t, y_{t-1}, \dots) = y_t + \sum_{j=0}^{\infty}(\sum_{i=j}^{\infty} b_i)\epsilon_{t-j} \tag{5.15}$$

In the same way, if we consider the optimal predictor $\hat{y}_{t-1}$ it is easily seen that:

$$\hat{y}_t - \hat{y}_{t-1} = (\sum_{i=0}^{\infty} b_i)\epsilon_t = b(1)\epsilon_t \tag{5.16}$$

therefore $\{\hat{y}_t\}$ is a pure random walk process unless $b(1) = 0$.

The processes $\hat{y}_t$ and $\Delta\hat{y}_t$ are called the long-run components of y_t and Δy_t.

In this way, the sequence $\{b_i; i = 0, 1, \dots\}$ represents the impulse-response function of the shock ϵ_t upon Δy_t, furthermore $h_{k,i} = \sum_{j=1}^{k+i} b_j$, $i = 0, 1, \dots$ is the impulse-response function of ϵ_t upon y_{t+k}. $b(1)$ is the impact of ϵ_t upon $y_{t+\infty}$, i.e. the limit of the impulse-response function as the horizon extends to infinity.

[14] See Hendry (1995), ch. 7, p. 231.

It can be easily shown that if y_t is a stationary linear process, then $b(1) = 0$, so that the long-run component of Δy_t vanishes. Consequently, unless $b(1) = 0$, $\hat{y}_t$ and hence y_t are non-stationary and includes a stochastic trend.

We can also decompose the process y_t in the long-run and short-run components. For this purpose we use the identity: $b(L)-b(1) = (1-L)b^*(L)$, where $b^*(L) = b_0^* + b_1^* L + b_2^* L^2 + \ldots$ and $b_0^* = -\sum_{i=1}^{\infty} b_i, b_1^* = -\sum_{i=2}^{\infty} b_i, \ldots$

Then, we have: $\Delta y_t \equiv b(L)\epsilon_t = b(1)\epsilon_t + (1-L)b^*(L)\epsilon_t$. This shows that Δy_t can be decomposed into the long-run component $b(1)\epsilon_t$ and the short-run component $(1-L)b^*(L)\epsilon_t$. In the same way we can write:

$$y_t = y_0 + \sum_{j=1}^{t} y_j = b(1)\sum_{j=1}^{t} \epsilon_j + b^*(L)\epsilon_t + y_0 - b^*(L)\epsilon_0 \tag{5.17}$$

where y_0 is the initial value for the initial time $t = 0$ considered in the conditional analysis of the processes. If we slightly modify the definition of the long-run component of y_t as follows: $\hat{y}_t = b(1)\sum_{j=1}^{t} \epsilon_j$, then we see that y_t is decomposed into the *long-run component* $b(1)\sum_{j=1}^{t} \epsilon_j$, the *short-run component* $b^*(L)\epsilon_t$ and the initial term $y_0 - b^*(L)\epsilon_0$.

The short-run component does not affect the predictive value of y for an infinite horizon period, so $b^*(L)\epsilon_t$; The only elements that matters are the long run component and the initial value (the value of the variable at time 0). Nevertheless, if the variable y is a stationary one, or I(0), the long run component must collapse to 0 for a very large horizon period. Therefore every deviation from the initial value that might occur within the horizon period is a temporary one, and the variable is expected to restore itself along the initial value. On the other hand, if we consider a non-stationary variable, say I(1), the long-run component is not zero and a trend is considered ($b(1) \neq 0$). In that case the variable is going to restore itself along with a different value from the initial one (it could follow a particular trend).

5.4.3 Spurious regression

The spurious regression problem is linked among others to the non-stationarity of the series. In the stationarity case, typically when the sample size increases, the law of large numbers applies and the sample variances and covariances converge to the population variances and covariances. This is a crucial property to show the consistency of estimators of the regression parameters. When the series are non-stationary the population variance and covariance are ill-defined because the series, as the time t increases, does not revert to their constant unconditional mean.

A simulation example analysed by Granger and Newbold (1974) gives evidence to this aspect. Consider two uncorrelated random walk processes:

$$y_t = y_{t-1} + u_t, u_t \approx iid(0, \sigma_u^2) \tag{5.18}$$

$$x_t = x_{t-1} + v_t, v_t \approx iid(0, \sigma_v^2) \tag{5.19}$$

where u_t and v_t are assumed to be serially uncorrelated as well as mutually uncorrelated. Furthermore, consider the regression model: $y_t = \beta_0 + \beta_1 x_t + \epsilon_t$.

Since y_t and x_t are uncorrelated random walk processes[15], we would expect the R^2 from this regression is near zero. But, if the two time series are growing over time, they can be correlated even if the increments in each series are uncorrelated. Simulations given by Granger and Newbold show a rejection rate of 76% when

[15] Granger and Newbold (1974) consider also a more general case when all the regressors are I(1) and not cointegrated.

testing the correct null hypothesis that $\beta_1 = 0$ in the regression using the conventional t-statistic. High levels of R^2 are often accompanied by highly autocorrelated residuals, as indicated by very low Durbin-Watson (*dw*) statistic. The inequality $R^2 > dw$ may be a useful rule of thumb suggesting that regressions should be run on the first differences of the variables. Phillips (1986) confirmed these empirical conclusions by finding that the OLS estimator does not converge in probability as the sample size increases, t- and F-statistics do not have well-defined asymptotic distributions and the *dw* statistic converges to zero.

Furthermore, Phillips (1986) and Entfor (1992) show that the result of a spurious regression depends on whether we consider random walk processes with drifts or with no drifts. If no drifts are considered into the two random walk processes, as in the previous model, the estimator $\hat{\beta}_1$ of parameter β_1 is convergent to a random variable; in the case of drift it converges to a constant. In both cases the estimator $\hat{\beta}_0$ of parameter β_0 is divergent.

To specify the model in a detrended form by using the log-difference operators is not necessarily a good choice because it can lead to further misspecifications in the model, especially when the non-stationary time series are characterized by long term relationships.

5.4.4 Autoregressive Distributed Lags Model (ADL)

ADL models or *autoregressive distributed lag models* were discussed by Hendry et al. (1984). A general ADL model with p regressors, m lags in y, and n lags in each of the p regressors is denoted by ADL $(m, n; p)$. It is given by:

$$y_t = \alpha_0 \sum_{i=1}^{m} \alpha_i y_{t-i} + \sum_{j=1}^{p} \sum_{i=0}^{n} \beta_{ji} x_{jt-i} + \epsilon_t \tag{5.20}$$

Consider the simplest ADL $(1, 1; 1)$ model:

$$y_t = \alpha_0 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t \tag{5.21}$$

where it is assumed that $\epsilon_t \sim \text{idd}(0, \sigma^2)$ and $|\alpha_1| < 1$.

If we define $y^* = E(y_t)$ and $x^* = E(x_t)$ for all t, we could find the ADL model in the long-run context.

Therefore, since $E(\epsilon_t) = 0$, $y^* = \alpha_0 + \alpha_1 y^* + \beta_0 x^* + \beta_1 x^*$ and hence $y^* = \frac{\alpha_0 + (\beta_0 + \beta_1) x^*}{(1-\alpha_1)} \equiv k_0 + k_1 x^*$ or

$$E(y_t) = k_0 + k_1 E(x_t) \tag{5.22}$$

where $k_0 = \frac{\alpha_0}{1-\alpha_1}$ and $k_1 = \frac{\beta_0 + \beta_1}{1-\alpha_1}$ is the long-run multiplier of y with respect to x.

5.4.5 Cointegration

Let us consider a set of non-stationary time series $x_{1,t}, x_{2,t} \ldots x_{n,t}$ which are integrated of order k so that $\Delta^k x_{i,t}$ is stationary, $i = 1, \ldots, n$. If there is a linear combination of the n series $x_{i,t} = b_1 x_{1,t} + b_2 x_{2,t} + \ldots + b_n x_{n,t} = w_t$ and w_t is integrated of order p lower than k, $w_t = I(p)$, $p < k$, then the series $x_{i,t}$ are called cointegrated of order p.

Particularly relevant in the macroeconomic context is the case where the series $x_{i,t}$ is integrated of order 1, so that: $x_{i,t} = I(1)$, $i = 1 \ldots n$. In such particular case, the linear combination above, if it exists, produces a w_t variable which is $I(0)$, then stationary. In economic terms this means that a set of n time series integrated of order 1, can converge to a steady-state equilibrium. Where a shock occurs, in the short term the system temporary diverges from equilibrium; whereas, after a finite time horizon, a new position in the steady-state equilibrium is reached following the same path defined by the linear combination above. The concept of

cointegration was introduced by Engle and Granger (1987) although if it was already present in Nelson and Plosser (1981) and Harvey (1981).

Going back to our simple regression model $y_t = \beta_0 + \beta_1 x_t + u_t$ and assuming that both y_t and x_t are $I(1)$, as said in 6.3, estimating the model in levels could lead to a strong misspecification and to misleading results due to the fact that both series are dominated by the trend. Nevertheless, if the series are also cointegrated so that: $b_1 y_t + b_2 x_t = w_t$, where $w_t = I(0)$, also the log-difference formulation of the regression will suffer of a big misspecification since it will neglect the presence of the long term equilibrium conditions represented by the equation above. By estimating the model in the first difference, the misspecification will appear by a unit root present in the estimated residuals $\hat{u}_t$ obtained by the OLS estimation. The presence of such a unit root can be tested by using either the augmented Dickey-Fuller or the Phillips-Perron tests on the series of the estimated residuals $\hat{u}_t$. In the case of the presence of cointegration, as suggested by Engle and Granger (1987), the best modelling strategy to be used is the error correction reparametrisation of the model.

5.4.6 Error Correction Model (ECM)

ECM models or *Error correction model* may be considered as a reparametrisation of the ADL ones. We can rewrite the simplest ADL (1,1;1) as follows:

$$y_t - y_{t-1} = \alpha_0 + (\alpha_1 - 1)y_{t-1} + \beta_0 x_t - \beta_0 x_{t-1} + \beta_1 x_{t-1} + \beta_0 x_{t-1} + \epsilon_t \tag{5.23}$$

Taking into account that $\beta_0(x_t - x_{t-1}) = \beta_0 \Delta x_1$, $\beta_0 x_{t-1} + \beta_1 x_{t-1} = (\beta_0 + \beta_1)x_{t-1}$ we have the following expression[16]:

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1)(y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \alpha_1} x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t \tag{5.24}$$

or

$$\Delta y_t = \alpha_0 + (\alpha_1 - 1)(y_{t-1} - k x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t \tag{5.25}$$

where $k = \frac{\beta_0 + \beta_1}{1 - \alpha_1}$.

Note that $(y_{t-1} - k x_{t-1})$ is the disequilibrium of the past period and k is the long-run multiplier of x_t with respect to y_t. Thus 5.25 is the ECM form implied by the ADL (1,1) model, with the ADL parameters (β_0, β_1, α_0,α_1) transformed in the ECM parameters (α_0, $(\alpha_1 - 1)$, β_0 and k). Engle and Granger (1987) show that, under a cointegration hypothesis between two time series (x_t and y_t), the ECM expressed in 5.25 is a suitable model in order to explain the dynamic of y_t in function of x_t.

The variation on y_t is given by the variation on x_t and the disequilibrium of the past period. At time t, this disequilibrium is increased or diminished by the coefficient $(\alpha_1 - 1)$ before affecting y_t.

Since in the 5.25 equation short-run and long-run relationships between time series are involved, the ECM model plays an important role in economic analysis and forecasting. The Engle-Granger two-steps procedure can be used in ECM estimation[17].

In the first step the OLS estimation is applied to the static relationship between x_t and y_t

$$y_t = k x_t + u_t \tag{5.26}$$

hence cointegration parameter[18] ($\hat{k}$) and residuals ($\hat{u}$) are estimated. If the series of residuals is proved to be stationary (for instance using the ADF test), then the ECM can be properly estimated. In particular, residuals

[16] See Bardsen (1989)

[17] The Engle-Granger two-step procedure assumes exogeneity for the explanatory variables. Outside this assumption the full information procedure is recommended (e.g. Johansen's cointegration procedure).

[18] The estimator $\hat{k}$ is superconsistent if both x_t and y_t are I(1) and cointegrated (see Stock (1987)).

of static regression become an explanatory variable in the ECM model as follows:

$$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + (\alpha_1 - 1)\hat{u}_{t-1} + \epsilon_t \tag{5.27}$$

Since all the elements that compose the 5.27 are stationary, OLS estimation has the usual optimality properties and the standard statistics can be applied for testing.

5.4.7 Regression with ARIMA errors: RegARIMA

Consider the RegARIMA[19] model in the matrix form:

$$y = X\beta + z \tag{5.28}$$

where $y = (y_1, \dots y_T)'$, X is a $T \times r$ matrix and $z = (z_1, \dots z_T)'$. In that model, the z_t follow an ARIMA process, $\phi(B)z_t = \theta(B)\epsilon_t$; while $\epsilon \sim \text{NID}(0, V_a)$. The variance of the process y_t is the variance/covariance matrix of an ARMA process.

The variance/covariance matrix of z[20] depends on AR and MA parameters and on the innovation variance (V_a). We also assume that y is normally distributed in order to find the GLS estimators of β and V_a by maximization likelihood function over parameters vectors θ and ϕ for the RegARIMA model.

Moreover, the GLS estimators are normally distributed even if inference is drawn separately, so the variance/covariance matrix of $\hat{\beta}$ can be estimated in function of the AR and MA estimated parameters. Hypothesis tests are then easily derived.

The researcher should choose the regressors, evaluate their qualitative grounds and identify a stochastic structure when exogenous regressors are present.

Firstly, regressors and regressands should be subjected to non-linear transformation, then with the OLS regression method we have the coefficients estimates ($\hat{\beta}_{OLS}$) and residuals estimates (the researcher does not know the model structure and OLS estimates are still consistent).

In order to apply the standard identification methodology, ACF and PACF must be tested; these tests should give important information about the stochastic structure and consistency estimation.

After knowing the ARIMA structure of errors and its intertemporal dependence, another non-linear transformation of errors should be done together with a GLS reestimation of regression coefficient ($\hat{\beta}_{GLS}$). An alternative approach based on Maximum Likelihood through the use of Kalman Filter is used in the FLASH software[21], see FLASH software (2002).

By forecasting, the researcher will find the future value of regressor, regressand and errors of the RegARIMA model.

Consider $y^* = (y_{T+1}, \dots, y_{T+k})'$, $X^* = (x_{T+1}, \dots, x_{T+k})'$ and $z^* = (z_{T+1}, \dots, z_{T+k})'$ the future variables. Moreover we denote $\sum_y = var(y)$, $\sum^* = var(y^*)$ the variance predictor, and $\sum_y^* = cov(y^*, y)$.

We suppose β and X^* as known and the stacked vector $(y, y^*)'$ normally distributed as follows

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(\begin{bmatrix} X\beta \\ X^*\beta \end{bmatrix}, \begin{bmatrix} \sum_y & (\sum_y^*)' \\ \sum_y^* & \sum^* \end{bmatrix} \right) \tag{5.29}$$

by applying multivariate normal distribution we may find predictors ($\hat{y}^* = E(y^*|y), Var(y^*|y)$) and the actual

[19] No lagged dependent variables and the error is assumed to follow an autoregressive moving average process (ARIMA).

[20] If z_t white noise is assumed, the best linear unbiased estimate of β would be OLS, on the other hand, if z_t are correlated, the Generalized Least Squared (GLS) estimator is considered.

[21] In particular the model is put in space state form and model parameters are estimated by ML (as described in Jones (1980)).

forecasting errors for the endogenous variable[22] $(y^* - \tilde{y}^*)$, the variance of the forecasting errors accounts for the uncertainty of the estimator for the regression coefficients of β.

5.4.8 ARMA models with exogenous inputs: ARMAX

An ARMAX model has the following structure:

$$\begin{aligned} \phi(L)y_t &= x_t'\beta + z_t \\ z_t &= \theta(L)a_t \end{aligned} \tag{5.30}$$

where a_t is a white noise, x_t is a vector of explanatory variables and the roots of $\phi(L)$ and $\theta(L)$ lie outside the unit circle.

If we consider the case in which $\theta(L) = 1$, ARMAX models are reduced to ARX models where the OLS estimations of parameters is consistent.

In the case of autocorrelated z_t terms methods based on instrumental variables should be considered (two stages least square and limited information maximum likelihood), in order to obtain a consistent estimation of model parameters $\phi_1, \phi_2, \ldots, \phi_p, \beta$. Nevertheless, researchers can choose more efficient methods based on ML[23]. In the case of exact ML model parameters estimators, space state model and Kalman filtering techniques should be used.

5.5 Rapid Estimates Based on Multivariate time series models

5.5.1 Vector Autoregressive (VAR) models

The Vector Autoregressive model, introduced by Sims (1980), is one of the most used tools for the econometric analysis of economic data. The rationale is to study a set of indicators, relating each of them, measured at a certain time t, to all the others and to all of their past realization up to a past time $t - p$. Two main events foreran the introduction of the VAR models: the large success of the monovariate time series models introduced by Box and Jenkins (1976) and the crisis of the traditional large simultaneous equations models inspired by the Cowles Commission work. The Box and Jenkins models had two revolutionary features: the possibility to describe dynamic behaviour and the apparent absence of an economic theory foundation, resulting in neutrality of the analysis. Their consensus was due mainly to their good performance in forecasting. The crisis of the traditional models brought to two critiques, the Lucas critique, based on the rational expectations theory, and the Sims critique, focusing on the absence of dynamic behaviour in the traditional models and the difficulty of a clear identification of the endogenous versus exogenous variables. Sims developed the VAR models merging the multiequation traditional approach and the time series approach. VAR models can then be seen as a multiple time series generalization of an Autoregressive model (AR).

A VAR model of order p is written as VAR(p) and is formalized as follows:

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{5.31}$$

where p is the time horizon and is, itself, a parameter to be determined; y_t is a vector $(n \times 1)$ of economic variables; c is a $(n \times 1)$ vector of intercepts;

ϕ_1, ϕ_2, ..., ϕ_p are $n \times n$ matrices of parameters and ϵ_t is the error term vector of dimension $(n \times 1)$.

[22] Through formulas we also find that $\tilde{y}^* = \hat{\beta} + \sum_y^* \sum_y^{-1} (y - X\hat{\beta})$

[23] An approximated starting point in the maximization procedure could be given by the 2SLS estimation of parameters. $\phi_1, \phi_2, \ldots, \phi_p, \beta$

Required hypothesis are $E(\epsilon_t) = 0$ and $\sum(\epsilon_t)$ is positive definite, which means ϵ_t follows a white noise stochastic process.

Despite the high level of complexity that VAR model can handle the VAR representation is very compact and synthetic permitting a simple and clear statement of the involved variables and their relationships. The determination of the p parameter, or the lag order, can be run, just like in the monovariate case, by using information criteria such as the AIC (Akaike (1981)) or the *final prediction error* (FPE) (see Lütkepohl (2006)). The estimation of a VAR model is conceptually simple. Since the same AR(p) model is adopted for every equation of the VAR model, and the stationarity is assumed, the OLS method for each equation can be adopted to estimate all the parameters. If the error process ϵ_t is normally distributed then the least squares estimators is equal to the maximum likelihood one. A very appealing feature of the VAR models is that they can easily take into consideration eventual non-stationary variables or even cointegrating relations among variables, introducing some parameters restrictions. The most general unrestricted VAR model allows, again, for an atheoretical approach, becoming the favorite tools of the neutrality fans opposing the structural models approach. This most appreciated feature of the general unrestricted model is also its major limit: the number of the parameters of an unrestricted model grows exponentially with the number of variables, making soon unfeasible the estimation. VAR models are data hungry, in the sense that they need a lot of observations to be estimated. This is the so called curse of dimensionality. It is possible to deal with this problem remaining within the atheoretical framework. One of the most important econometric methods of this type is the so called London School of Economic (LSE) approach, for which the initial model is the most general as possible and then the data are left to speak, testing for significance of the parameters and getting to the final model by an error and try iterative process. The best description of this approach is in Hendry (1995). While the LSE approach deal with the complexity introducing parameters restrictions based on a massive testing process other approaches can be applied, for example reintroducing restrictions theoretically justified or reducing the number of the variables by aggregating a subset of them with a principal component, or similar, method. The possibility of implementing cointegration relations allows for modelling and testing theoretical economic relationships, bridging to Structural VAR (SVAR) model (Sims 1986). Those kinds of models perform well in forecasting and in a priori evaluation of policies. Their utility in the construction of rapid estimates is straightforward.

5.5.2 Vector Error Correction Models (VECM)

The VAR model has always a different representation called Vector Error Correction Model (VECM). This representation can be a useful tool when non-stationary cointegrated series are treated. Considering the generic VAR(p) representation:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t, \tag{5.32}$$

where y_t is the vector $(n \times 1)$, the ϕ_i are $(n \times n)$ parameters matrices and the ϵ_t is the error term vector of dimension $(n \times 1)$. Applying some algebra the VEC representation becomes:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \hat{\phi} \Delta y_{t-i} + \epsilon_t, \tag{5.33}$$

where Π and $\hat{\phi}$ are respectively:

$$\begin{aligned} \Pi &= -(I - \phi_1 - \ldots - \phi_p) \\ \hat{\phi} &= -\sum_{i=j+1}^{p} \phi_i, j = 1, \ldots, (p-1) \end{aligned} \tag{5.34}$$

The advantage of the VECM is that they are subject to an easy and clear interpretation based on the properties of the Π matrix:

- If Π has positive, but not full, rank: $0 < \text{Rank}(\Pi) = m < n$ then there are m cointegrating relations among the n series. In this case it is possible to factorize the Π matrix as follows: $\Pi = \alpha\beta'$ where β contains the m cointegrating vectors and α contains the m adjustments vectors;
- If $\Pi = 0$ then there is no cointegration, non-stationarity of the y_t components vanishes with differentiating them singularly;
- If Π has full rank, n, then each components of y_t is already stationary.

The second and the third cases are not interesting while the first case allows for interpreting the Πy_{t-1} part of formula 5.33 as the long run relation: equilibrium is on $\Pi y_{t-1} = 0$. If the long run relation does not hold then there will be an adjustment towards the long run equilibrium, at a speed determined by the parameter contained in α, this part is also called error correction term since corrects deviations from long run equilibrium, the errors.

$\sum_{i=1}^{p-1} \hat{\phi}\Delta y_{t-i}$ will then be the short term part of the formula 5.33 containing adjustments to short term disequilibrium and ϵ_t will represent the shocks bringing the system out of both, short and long term, equilibria. This ensure that the variables evolve along time following m common paths.

As a practical example, we take into account a transformation of a bivariate ($n = 2$) VAR(1) in a VECM form. Let us consider the following bivariate VAR(1) without intercept:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{5.35}$$

where $y_t' = \begin{pmatrix} y_{1,t} & y_{2,t} \end{pmatrix}$.

If we apply the VECM transformation we obtain:

$$y_t = \Pi_{t-1} + \phi\delta y_{t-1} + \epsilon_t, \tag{5.36}$$

where $E_t' = \begin{bmatrix} \epsilon_{1t} & \epsilon_{2t} \end{bmatrix}$. The matrix Π, the long-run matrix, is given by:

$$\begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \tag{5.37}$$

and has reduced rank if:

$$(a_{11} - 1)(a_{22} - 1) - a_{21}a_{12} = 0. \tag{5.38}$$

Under an arbitrary normalization, we can have the following cointegration vector:

$$\beta' = [a_{11} - 1 \qquad a_{12}] \tag{5.39}$$

Imposing 5.38 and 5.39 the factorization of Π become:

$$\Pi = \alpha\beta' = \begin{bmatrix} 1 \\ a_{21}/(a_{11} - 1) \end{bmatrix} [a_{11} - 1 \qquad a_{12}] \tag{5.40}$$

The VECM associated with this factorization is:

$$\begin{cases} \Delta y_{1,t} = (a_{11} - 1)[y_{1,t-1} - \lambda y_{2,t-1}] + \epsilon_{1,t} \\ \Delta y_{2,t} = a_{21}[y_{1,t-1} - \lambda y_{2,t-1}] + \epsilon_{2,t} \end{cases} \tag{5.41}$$

where $\lambda = a_{12}/(a_{11} - 1)$ is the long-run coefficient. Note that in this example we can have only one cointegration relation ($m = 1$) and the $y_{1,t}$ and $y_{2,t}$ processes share only one common trend. As already seen

we cannot have two cointegrating relationships: two stationary linear combinations of $y_{1,t}$ and $y_{2,t}$ imply the stationarity of both time series themselves. Furthermore, rank $r = 0$ indicates the absence of cointegrating relationships and both $y_{1,t}$ and $y_{2,t}$ are integrated but non-cointegrated processes: for example, if $a_{11} = a_{22} = 1$ and $a_{12} = a_{21} = 0$, then $y_{1,t}$ and $y_{2,t}$ are two random walks.

A single deviation from the long-run equilibrium could be corrected by repeated adjustments. In our particular example the adjustments are represented by the error correction terms $[y_{1,t-1} - \lambda y_{2,t-1}]$ and both the adjustment coefficients $(a_{11} - 1)$ and a_{22} determine the speed of adjustment of their respective endogenous variables towards the equilibrium.

5.5.3 Global VAR (GVAR)

The increasing degree of interdependency at geographical and sectorial level as well as between real macroeconomic and financial sectors is calling for a new modeling framework for incorporating such effects. The main problem related with the increasing number of variables and their interaction effects is the so called curse of dimensionality. The number of parameters to be estimated in a VAR grows exponentially with the number of the variables making soon unfeasible the estimation.

Pesaran et al. (2004) gave an answer to this need by developing the so-called global vector auto regression models (GVAR). This family of models was then extended in Pesaran et al. (2004) and its forecasting abilities were analyzed in Pesaran et al. (2008).

The GVAR model is composed of a large number of unit-specific models, actually each sector/country/region is modeled by means of one vector error correction model and cross-dependency effects are allowed so that internal variables depends on external variables affecting the unit from outside the borders. In practice main internal macroeconomic variables, such the GDP of a given country can depend on global observed variables, such as international energy or food prices, and on foreign specific variables corresponding to the related internal one, such as the foreign GDP. Country- specific foreign variables are built as weighted averages of other countries' variables, usually using trade weights, and they can be interpreted also as proxies for common unobserved factors.

In the GVAR framework, all individual VECM are estimated in a consistent way, each separately but conditional on the foreign variables which are assumed to be weakly exogenous. Weak exogeneity assumption must be tested, but is a reasonable assumption considering the small size of each single country economy with respect the global one. This assumption is the crucial one for the model since it allows estimating the specific models individually, dealing with complexity without falling into the curse of dimensionality.

Short run dynamic is usually left unconstrained while the long run properties can be linked to market arbitrage and to specific stock-flow equilibria, allowing for long term consistency with the theory and short run consistency with the data. GVARs allow for taking into account several channels of shock transmissions. The observed interdependencies can be determined by some global shocks or can originate by specific country or sectorial shocks. As mentioned by Ericsson and Reisman (2012) the GVAR models have some attractive features which can be summarized as follows: multiple potential channels for the international transmission of macroeconomic and financial shocks, a standardized economically appealing choice of variables for each examined country or region, systematic treatment of long-run properties through cointegration analysis, and flexible dynamic specification through vector error correction modelling.

In order to formalize the GVAR model let's consider a simple model describing a global economy with $(N+1)$ countries: each country economy has an idiosyncratic VAR representation. For simplicity an order 2 VAR with no global variables, for the generic country i, is considered:

$$y_{it} = a_{i0} + a_{i1}t + \phi_{i1}y_{i,t-1} + \phi_{i2}y_{i,t-2} + \Lambda_{i0}y^*_{i,t} + \Lambda_{i1}y^*_{i,t-1} + \Lambda_{i2}y^*_{i,t-2} + \eta_{it}, \tag{5.42}$$

where:

- $i = 0, 1, \cdots, N$ (country 0 serves as reference[24]);
- y_{it} is a vector $(ki, 1)$ of country-specific macroeconomic variables: GDP, inflation, interest rates, exchange rates, etc.;
- y^*_{it} is the vector $(k^*i, 1)$ of corresponding foreign variables;
- $a_{i0} + a_{i1}t$ is vector of deterministic trends; ϕ's and Λ's are matrices of parameters;
- η_{it} is a vector serially uncorrelated and cross-sectional weakly dependent process.

Foreign-specific variables are weighted averages of the corresponding domestic variables for all the countries with exogenous country-specific weights:

$$y^*_{it} = \sum_{j=0}^{N} w_{ij} y_{jt} \tag{5.43}$$

with $j = 0, 1, \ldots, N$; $w_{jj} = 0$ for all j, and $\sum_{j=0}^{N} w_{ij} = 1$. Weights can be derived by international trade volumes or any other variable representing the relevance of country j in the economy of country i.

The VECM representation of the individual VAR will be:

$$\Delta y_{it} = c_{i0} + \alpha_i \beta_i' (z_{i,t-1} - \lambda_i (t-1)) + \Lambda_{i0} \Delta y^*_{i,t} + \Gamma_i \Delta z_{i,t-1} + \eta it \tag{5.44}$$

where:

- $z_{i,t} = (y_{it}', y_{it}^{*'})'$;
- α_i is a $k_i \times r_i$ matrix of rank r_i of adjustment coefficients;
- β_i is a $(k_i + k_i^*) \times r_i$ matrix of rank r_i of cointegration parameters;
- $z_{i,t-1} - \lambda_i (t-1)$ is the long run equilibrium relation;
- $\Lambda_{i0} \Delta y^*_{i,t}$ is the contemporaneous relation between internal and external variables;
- $\Gamma_i \Delta z_{i,t-1}$ is the short term equilibrium relation.

The long run equilibrium allows for cointegration within y_{it} and between y_{it} and y^*_{it}, throughout y^*_{it} also cointegration between y_{it} and y_{it}, with $j \neq i$, is allowed. A vector of global variables, exogenous or weakly endogenous, can be easily added to the model.

The weak exogeneity of the y^*_{it} and of the eventual global variables, allow for country specific estimation of the model, the global model will be then estimated recursively, recalculating the foreign-specific and the global variables for each country.

The relevance of GVAR for constructing rapid estimates can be really high, both for individual countries and economic areas and regions such as the Euro area. A wide review of the literature, as well as the latest development in the GVAR modelling, is presented in Chudik and Pesaran (2014) or in di Mauro and Pesaran (2013). A quasi natural extension of the GVAR modelling is the panel VAR (see Canova and Ciccarelli (2013)). Nevertheless, at this stage, we consider this extension still not completely assessed and quite too complex to be proposed within this handbook.

[24] or numeraire

5.6 Rapid estimates models based on data available at different frequencies

In some cases the indicator series can be available at a higher frequency than the target variable. For example, when nowcasting the quarterly GDP, we often use indicators such as monthly industrial production index, retail trade deflated turnover, tendency surveys, etc., which are usually available at frequency higher than quarterly. On the other hand, in official statistics, it happens that the higher reliable estimate of an indicator is available at lower frequency. This is often the case for National Accounts where annual aggregates constitute the most reliable estimates being based on a quite exhaustive set of surveys, possibly integrated by other sources of information such as administrative ones. In such a case, quarterly national accounts can be derived by distribution (temporal disaggregation methods) of annual ones using relative quarterly proxy information. Both cases require the use of some statistical specific techniques dealing with models available at different time frequencies. In the first case, the easiest way to deal with the mixed frequency problem is to transform data from high-frequency to low-frequency by time aggregation procedures.

Consider y_{th} a high-frequency series which must be transformed into a low-frequency one y_{tl}. Users must apply a particular transformation procedure in order to get this result. In general, a transformation matrix B is used, so data should be converted following this formula: $y_{tl} = By_{th}$. Aggregation could be made by *mean* (for the case of index series, averaging the corresponding periods) or by *sum* (for the case of flow series, summing up the corresponding periods) or by *interpolation* (for the case of stock first and stock last series, imputing the value of either of the first quarter to the annual or the value of the last quarter to the annual; think about the case of the population that is measured at the beginning of the year of the stocks series); The form of the matrix B should reflect data characteristics with particular attention to the distinction between flow and stocks and index.

Choosing the wrong aggregation constraint might produce misleading results, so that particular carefulness must be taken when dealing with data conversion. Converting data from high to low frequency can imply a loss of information so that it could be advisable to use more sophisticated techniques dealing with multi-frequency data modelling such as bridge equations and mixed-data sampling (MIDAS).

On the other hand, if the high frequency data is characterised by high degree of volatility, converting it into low frequency one can mitigate the importance of this component so that the model specification can be improved. In the second case, the easiest way to disaggregate low frequency data into high frequency one could be to assume an uniform distribution criterion (e.g. dividing annual data by four to obtain quarterly one). This would be a not an acceptable final solution (it should rather be used for as naïve scenario for benchmarking the final results) and the use of more sophisticated statistical techniques such as temporal disaggregation ones is required. Once again, by using such techniques, it will be crucial to use a correctly specified aggregation matrix B, reflecting the data typology and the aggregation constraint across frequencies.

Finally, since flash estimates must be computed using a methodology reflecting as close as possible the one used in the regular production process, if regular estimates are produced by using temporal disaggregation or mixed- frequency modelling techniques, flash estimates should also use them. In this section, temporal disaggregation and mixed-frequency modelling techniques will be briefly introduced in sub sections 5.6.1 and 5.6.2 respectively, while they will be deeper discussed in chapter 8 and chapter 9.

5.6.1 Temporal disaggregation methods

Temporal disaggregation techniques include a wide range of techniques stemming from purely mathematical ones (smoothing) to more sophisticated ones based on econometric or time series models. They can deal with one single indicator (univariate methods) or simultaneously with a number of indicators related by some aggregation or accounting constraints (multivariate methods). An interesting review of the literature is proposed by Di Fonzo (1987) and in the Eurostat Handbook on Quarterly National Accounts (Eurostat (1999))

and Poissonnier et al. (2013) (chapter 5 of the Eurostat Handbook on Quarterly National Accounts, 2013 Edition). Temporal disaggregation methods can be broadly classified as follows:

1. When there is a lack of information, smoothing methods ensures the possibility of obtaining both interpolation estimates for the quarterly breakdown and extrapolations for now-casting during the year. They are usually based on mathematical techniques and do not allow, in general, direct extrapolation (e.g. Boot et al. (1967)). When using methods of this type, the annual data need to be inspected before smoothing. In cases where the annual series has null values, then unsatisfactory quarterly estimates, including negative values, may be produced, and these should be amended manually. If there are consecutive annual null values, then the smoothing method may generate a quarterly seasonal pattern in those years that will require amendment. The major difficulty with smoothing methods occurs in the latest year, when there are no annual data.
2. The two steps adjustment methods divide the process of estimation in two parts. The first step consists of a preliminary estimate of the aggregate that does not fulfill the annual constraints. The second step consists in distributing the discrepancy according to appropriate criteria between annual data and preliminary estimates (e.g. Bassie (1958) and Denton (1971)).
3. Time series methods can be used both in situations with lack of information where, allowing the extrapolation, they represent a more sophisticated statistical smoothing method, and as adjustment-optimal methods (e.g. Stram and Wei (1986); Wei and Stram (1990) and Al-Osh (1989)). Time series methods can readily produce quarterly estimates beyond the latest annual benchmark, but if exogenous information is available for the latest year, it is best to use it make a formal estimate. It is also possible to expand the model to incorporate information derived from quarterly related series in order to guide the infra-annual path.
4. Regression based methods merge the steps of preliminary estimation and adjustment in one statistically optimal procedure using all the available information in the context of a statistical regression model which involves annual information and quarterly related information. (e.g. Chow and Lin (1971); Bournay and Laroque (1979); Fernandez (1981); Litterman (1983); Di Fonzo (1987))
5. The dynamic regression model based methods extend the optimal approach by incorporating dynamic elements that permit consideration at the same time of the short-term influences derived from related series and of the effects of the recent past history of the aggregate to be estimated. (e.g. Salazar et al. (1997); Mitchell et al. (2005); Santos Silva and Cardoso (2001)).
6. Finally multivariate methods take into account the multivariate dimension of national accounts introducing contemporaneous accounting constraint in the estimation step in order to obtain estimates of national accounts aggregates which are coherent both temporally and in the accounting sense.

Within the ESS regular production process of official statistics, two steps adjustment methods, regression based methods and multivariate methods are the most commonly used, especially to produce macroeconomic indicators such as quarterly national accounts. Nevertheless, the first and the third ones are mainly used to reconciliate preliminary estimates of high frequency indicators with low ones over the time and to ensure simultaneously the fulfilment of temporal and accounting constraints. For this reason, their relevance in producing rapid estimates is relatively low. By contrast, regression based methods are a very useful tool to estimate high frequency indicators starting from low frequency ones and using all the available statistical information in a statistically robust and sound context, represented by the linear regression model. For this reason, they can be particularly useful when building up rapid estimates and in particular flash ones.

The Chow - Lin method and its variants

Chow and Lin (1971) proposed an innovative method to estimate monthly unobserved data starting from quarterly observed one, based on the linear regression model. They worked out a least-square optimal solution on the basis that a linear regression model involving the variable y and one or more related quarterly

indicators x can be estimated using the annual data for the y variable and the annualized quarterly data for the indicator(s). The Chow and Lin method has a several desirable features:

- It is very general in that it produces identical solutions to other methods, such as the BFL and the Denton difference benchmarking procedure, under certain assumptions.
- It is possible to evaluate the precision of the estimates by means of all the standard regression analysis tools. In particular:
  - compute confidence intervals for the estimates according to suitable distribution hypotheses;
  - assess the quality of the quarterly indicators;
  - perform econometric analyses that take into account the "noisy"' nature of the data.
- The concepts of the Chow and Lin method can be easily extended to the multivariate case and to more complex models.

Below, we propose a generic version of the method aiming to estimate an unobserved aggregate at high frequency l, based on the observed aggregates at low frequency h and a set of related information available both at low and high frequencies level. Our description mainly follows the one proposed in the Eurostat Handbook on quarterly national accounts (2013, edition). Let consider the following univariate temporal disaggregation problem: given the $(N \times 1)$ vector y_l of low frequency observations of an economic aggregate and given the $(n \times k)$ matrix X_h of high frequency observations on k related series, we wish to estimate the $(n \times 1)$ vector of unobserved high frequency observations y_h. The dimension of the $(n \times 1)$ vector is p times the one of the low frequency vector $(N \times 1)$, where p is the aggregation order, so that if the low frequency data is annual and the high frequency is quarterly, $p = 4$, if the low frequency data is quarterly and the high frequency is monthly, $p = 3$, etc. We suppose that the following linear regression model holds:

$$y_h = X_h\beta + u_h, \tag{5.45}$$

where β is a vector $(k \times 1)$ of unknown parameters and u_h is a vector of $(n \times 1)$ random disturbances such that:

$$E(u_h|X_h) = 0 \qquad E(u_h u_h'|X_h) = V_h. \tag{5.46}$$

The best linear unbiased estimator of y_h consistent with the temporal aggregation constraints $By_h = y_l$ is given by:

$$\hat{y}_h = X_h\hat{\beta} + V_h B' V_l^{-1} + \hat{u}_l, \tag{5.47}$$

where $\hat{u}_l = y_l - X_l\hat{\beta}$ and $\hat{\beta} = (X_l' V_l^{-1} X_l)^{-1} X_l' V_l^{-1} y_l$

It is observed that:

1. $\hat{\beta}$ is the generalised least squares estimate of β in the low frequency observed model:

   $$By_h = BX_h\beta + Bu_h, \tag{5.48}$$

   i.e. $y_l = X_l\beta + u_l$, where the $(N \times 1)$ vector of the low frequency aggregated random disturbances $u_l = Bu_h$ is such that:

   $$E(u_l|X_l) = 0 \qquad E(u_l u_l'|X_l) = V_l = BV_h B' \tag{5.49}$$

2. The aggregation constraints are satisfied:

   $$B\hat{y_h} = BX_h\beta + BV_h B' V_l^{-1}\hat{u}_l = y_l, \tag{5.50}$$

3. The high frequency estimates can be viewed as the sum of two components: a systematic one, given

by $X_h\hat{\beta}$, and an adjustment term, given by $L\hat{u}_l$, where L is the $(n \times N)$ smoothing matrix:

$$L = V_h B' V_l^{-1}, \tag{5.51}$$

A very important feature of this estimation approach is the capability to get an evaluation of the precision of the estimates by means of the estimation errors covariance matrix (Bournay and Laroque (1979)):

$$E[(\hat{y_h} - y_h)(\hat{y_h} - y_h)'] = (I_n - LB)V + (X - XL_l)(X_l V_l^{-1} X_l^1)(X - LX_l)' \tag{5.52}$$

This matrix depends on two components: the former is only related to B and V, the latter is a systematic one and rises with $(X - LX_l)$.

The estimation errors covariance matrix can be used to:

1. compute confidence intervals for $\hat{y_h}$ under the hypothesis $u \sim N(0, V)$;
2. assess reliability indicators for the estimates like $\frac{100 * \hat{\sigma}_{\hat{y}}}{\hat{y}}$ (Van der Ploeg (1985), p.9);
3. make econometric estimations taking into account the "noisy"' nature of the data.

The optimal estimation approach offers also a natural and coherent solution to the extrapolation problem.

In most practical problems matrix V is unknown and must, therefore, be estimated according to suitable assumptions on u. The stochastic models usually considered are the following:

1. AR(1) model (Bournay and Laroque (1979))

$$u_i = \rho u_{i-1} + \epsilon_i, |\rho| < 1. \tag{5.53}$$

2. Random walk model (Fernandez (1981)):

$$u_i = u_{i-1} + \epsilon_i, u_0 = 0. \tag{5.54}$$

3. Random walk-Markov model (Litterman (1983), Di Fonzo (1987)):

$$\begin{array}{rcl} u_i & = & u_{i-1} + e_i \\ e_i & = & \mu e_{i-1} + \epsilon_i, \end{array} \quad |\mu| < 1, u_0 = e_0 = 0. \tag{5.55}$$

The regression model used in the Chow and Lin method has no dynamic components, i.e. no lags or leads of y and x. The temporal disaggregation and interpolation processes are based only on the dynamic of the indicator. Substantial developments have been made to include dynamic features (Santos Silva and Cardoso (2001)) or to imbed them in the more general framework of the Kalman filter (Harvey and Pierse (1984); Moauro and Savio (2005) or Mazzi et al. (2005)). An alternative way to deal with temporal disaggregation is represented by the state-space representation, where all static and dynamic, univariate and multivariate models can be easily treated in a computationally efficient way. This approach will be illustrated in detail in chapter 8.

5.6.2 Mixed frequency models

The macroeconomic dataset on which forecasting exercises are built up are often unbalanced. The main reasons for this is, from one hand, the fact that target variables to be forecasted and its main indicators are available at different frequencies (usually indicators are available at higher frequencies than the target variable), from the other hand, the so-called ragged-edge problem due to the different publication delays of variables. The two factors determine the presence of mixed observations in the dataset. As mentioned

at the beginning of this section, the easiest solution to this problem is to convert the high frequency data into the same frequency of the target variable but this approach cannot be considered at an optimal one. It determines important loss of useful information and it can also determine mis-specification in the models used for forecasting. In the literature, several models have been proposed to deal with the presence of mixed-frequency data and ragged-edge effects in order to avoid the loss of relevant information available at higher frequency. Among those models, we mention: the Bridge regression (Baffigi et al. (2004) and Diron (2008)), the mixed data sampling models (Ghysels et al. (2004)), the mixed frequency VAR (Mariano and Murasawa (2010) and Schorfheide and Song (2011)), the mixed frequency factor models (Mariano and Murasawa (2003); Frale et al. (2010); Frale et al. (2011); Giannone et al. (2008) and Bańbura and Rünstler (2011)). A very detailed overview of such models is provided in chapter 9.

At this stage, it is important to notice that the proposed models have relevant advantages and drawbacks. In particular, Bridge models are linear and very easy to be computed but they don't deal with the ragged-edge phenomenon because they require that the high frequency data cover entirely the last period of the low frequency target variable. To overcome this limitation, it is possible to use forecasting techniques to fill the gaps but this can also produce biased results. On the other hand, the MIDAS models, despite their appealing features, can be computationally very complex and suffer of convergence problems due to their intrinsic non-linearity. This problem can be partially solved when using their unrestricted version as proposed by Foroni et al. (2012).

5.7 Dealing with competing rapid estimate models

When developing a system of rapid estimates for a given variable y_t it is common practice to compare a number of models from the simplest to the more complex ones. Furthermore, different kinds of data sources can also be used. The final expectation is that there is one model which clearly overcomes the others in terms of forecasting performance. Unfortunately, this is not often the case because several competing models can show similar forecasting abilities over the simulation period. Furthermore, some of them can perform better in a particular phase of the business cycle such as the expansionary one and others more in the recessionary one, etc.

Let consider the case in which we are trying to identify the best performing model in the one-step-ahead forecasting exercise for the variable y_t $T = 1, \ldots, T$. The comparison among alternative predictors is carried out over the period $T + 1$, $T + r$ while the competing models are estimated over the period $t = 1, \ldots, T$. The results of this exercise is a series of p competing series of one-step-ahead forecasts of y_j $\hat{y}_{i,j}$, where $i = 1, \ldots, p$ and $j = T + 1, \ldots, T + r$. At this time the strategy can be to use different criteria and tests to choose the most suitable predictor (see 5.9.10). An alternative could be to use all the information available from different competing predictors to build up a new one.

5.7.1 Combining Forecasts

The idea that combining various forecasts can lead to a better forecast in terms of the forecasting error was advocated by Crane and Crotty (1967) and Zarnowitz (1967) but it was the paper from Granger and Bates (1969) which is considered the seminal work in this field. The main idea is that, given a set of competing, independent forecasts of a given variable y_t, $t = 1, \ldots, T$, constructed over the forecasted horizon $T + 1, \ldots, T+r$, $\hat{y}_{i,j}$, where $i = 1 \ldots k$ and $j = T+1, \ldots, T+r$, a better forecast can be obtained by combining them so that

$$\hat{y}_{c,j} = \frac{\sum_{i=1}^{k} h_i \hat{y}_{i,j}}{\sum_{i=1}^{k} h_i},$$

for all $j = T+1, \ldots, T+r$, where h_i are the weights.

Bates and Granger proposed and compared several weighting schemes. The paper of Bates and Granger generated a very large theoretical and empirical literature as well as a big debate among various schools of thinking on the usefulness of the combining forecasting exercise. An interesting review of the literature is proposed by Clemen (1989). The relevance of combining forecasting exercise is widely discussed in chapter 12, while a more empirical approach in looking at forecasting combination is provided in chapter 13 of this handbook.

It is useful to note that the effect of the forecasting combination is the reduction of forecasting errors but this approach does not correct any existing bias in the components so that it is better to combine unbiased forecasts. It is also interesting to note that the combining exercise is meaningful if none of the competing forecasts encompass the others (see 5.9.11). Finally, the forecasting combination techniques have been largely used, also to derive density forecasts in opposition to the traditional point forecasts. This use of combining forecasting techniques is presented in Mazzi, Mitchell and Montana (2013).

5.8 Rapid Estimates based on available survey data

Of course, one natural but difficult way to construct a rapid estimate would be to speed up the production process. Most of short term economic indicators are built from survey data. The IPI for example is derived from a survey on the business industrial production. Getting enterprise answers quicker would give an earlier estimate of the IPI. The problem is that it takes time to collect the information and, if you want to run an early estimate, you will have non-responses and missing data.

5.8.1 Non-parametric models to assign values to the aggregates of the non-respondent units

We could distinguish between two types of missing data:

- total or unit non-response: for some elements selected for the observation (statistical units) no data are available at all.
- item non-response: for some elements selected for the observation (statistical units) only some data are available.

Since the missing values of some elements in the aggregates are considered as a total lack of data for such elements, we stress to the importance of the unit non-response problem. In order to deal with this problem, the non-parametric approach based on the nearest neighbour imputation (Chen and Shao (2000)) can be used. This method, named Nearest Neighbour Imputation, involves defining a distance measure in function of the auxiliary variables. The observed unit with the smallest distance to the non-respondent unit is identified and its value is substituted for the missing item according to the variable of concern. Consider an auxiliary variable X and the variable of interest Y. A bivariate sample could be constructed in the following way: $(x_1, y_1), \ldots, (x_n, y_n)$ where r of the n y-values are observed (respondent) and the rest $m = n - r$ y-values are missing (non-respondents).

Therefore the values $y_{r+1}, y_{r+2}, \ldots, y_n$ are missing. Any missing value $y_j (r+1 \leq j \leq n)$ can be replaced by y_i, where $1 \leq i \leq r$ and i is the nearest neighbour of j measured by the auxiliary variable. Formally this method satisfies this minimization problem:

$$|x_i - x_j| = \text{Min}_{1 \leq l \leq r} |x_l - x_j| \tag{5.56}$$

Although this non-parametric approach is deterministic, it estimates distribution correctly. In that context the observed values are sufficient to apply this method but sometimes the variance of the estimator may be

inflated. In fact some values could be used several times if more missing values occur and others may be not used at all. Restriction on observed value usage for non-respondent units may reduce the variance problem (Kalton (1983)).

5.8.2 Super population models: based on re-weighting techniques that correct weights assigned to timeliness respondents, so that they can also represent non-timeliness units in a suitable way

Weighting is a common form of unit non-response adjustment in sample survey. As in Little (1986), a common preliminary is to classify respondent and non-respondent into an adjustment cell, formed on the basis of information recorded for all units in the sample. Two adjustment procedures were followed by Little: imputation of cell mean (at all the non-response units are assigned the sample mean value) and weighting by reciprocal cell response rate (therefore, the respondent unit represents the non-respondents ones in a suitable way). The variance of population estimates is strictly affected by the method used to assign respondent values to non-respondent units.

The second case, called also quasi-randomization approach, assumes that every elements in the sample has a probability $\phi_i > 0$ $(i = 1, 2, \ldots, N)$ of responding if sampled (Oh and Scheuren (1983)). This approach would improve the not applicable MCAR[25] assumption for non-respondent units. As in the nearest neighbour imputation approach introduced by Chen and Shao, auxiliary variable should be introduced in order to obtain available values for both respondent and non-respondent units. Since each cell would be redefined in function of these new auxiliary variables, in our case, the weight of the respondent unit should be inflated by the reciprocal response rate in the cell.

The population estimation of the mean is unbiased if the $\phi_{ci} = \phi_c$ (MAR assumption[26]) where i denotes the element i in the cell c, and $\bar{\phi}_c$ is the average response probability in the cell c.

Auxiliary variables are also used by Rizzo et al. (1996). They have developed some weighting methods for non-response adjustment using the know characteristics of respondent and nonrespondent units (auxiliary variables). Firstly, they have carried out a screening strategy in order to detect which auxiliary variables can be the right candidates in the weighting adjustment procedure (by eliminating the ones which have little association with nonrespondent units). Subsequently, a convenient combination of these selected items is used in order to predict the units response status. The relationship between units response status and selected auxiliary variables has to be done by a logistic regression[27]. This approach is adopted to choose the best predictable variables to insert into the logistic regression model for having good predictors of units nonresponse. By knowing how statistical units are behaving respect to the selected auxiliary variables, we can collect those which have the same behaviour with respect to the selected items into the adjustment cells and calculated the predicted nonrespondent rate for that cell taking into consideration the logistic parameters estimation. Although the variance in the estimates is less inflated, the model provides a general result for the status prediction of a particular unit and it does not reflect the outcome in a specific sample. Furthermore, Rizzo Kalton and Brick proposed other alternative models, for instance the classification trees[28]. It has to detect the best strategy in adjustment cells selection. The adjustment cell are defined according to the identification of such predictor variables that have the largest discrimination between its responses categories and nonresponse rate. Giommi (1987) introduced a non-parametric approach in order to estimate the response probability in a sample survey. The probability of timeliness respondent can be modified by non-parametric techniques in order to represent non-timeliness units in a suitable way.

[25] The Missing Completely at Random (MCAR) approach is used when the probability of an observation being missing does not depend on observed or unobserved measurement. $\phi_i = \phi$ constant

[26] The Missing at Random (MAR) approach is used when, given the observed data, the missingness mechanism does not depend on the observed data.

[27] These selected variables have an explanatory power that is very significant with respect to the global specification of the model (using the Wald statistic).

[28] The screening of auxiliary variables are still used by applying this method.

Let us consider a finite population U of N size $U = (1, 2, \ldots, k, \ldots N)$ and, from that population, a sample of n units is drawn. The probability to draw a unit so, to include it into the sample, is given by $\pi_k > 0, \forall k$ and $\pi_{kl} > 0$ for $\forall k \neq l$ if two units are considered into the extraction.

Moreover, consider a particular set of variables, say $(y_1, \ldots, y_q, \ldots, y_Q)$, where Q denotes characteristics of the population under analysis. At each unit sampled we attach a probability $\varphi_k > 0$ to participate at a survey; if we consider two units the probability becomes $\varphi_{kl} > 0$ for $k \neq l$, otherwise if $k = l$ it reduces to $\varphi_k > 0$.

Although a particular unit participates to the survey, it could or could not give any information about a specific variable y_q. So, we denote with $\psi_{qk} > 0$ the probability to respond at a specific characteristic when a unit k is selected and participate at a survey. In case of two units (k and l) that probability becomes $\psi_{qkl} > 0$.

Therefore, $\pi_k \Theta_{qk}$, where $\Theta_{qk} = \varphi_k \psi_{qk}$, is used for denoting the probability that the k-th unit is respondent if it is selected and participate at a survey (if two units are considered, probability becomes $\pi_{kl}\Theta_{qkl}$, where $\Theta_{qkl} = \varphi_k \psi_{qkl}$).

We assume that if the variable y_q is not available for all the units sampled, an auxiliary one x_q can be considered (y_q and x_q are strongly correlated). This variable is known for every population units.

Niyonsenga (1994) and Giommi (1987) have argued that φ_k and ψ_{qk}, and of course, $\Theta_{qk} = \varphi_k \psi_{qk}$, are response probabilities that should be estimated via non-parametric estimation.

The probability φ_k and ψ_{qk} are function of the two auxiliary variables say z_k for φ_k and x_{qk} for ψ_{qk}. For simplicity, let us consider $z_k = x_{qk} = x_k$[29] for all q and for every unit sampled (s is the generic unit sampled) and x_s is the auxiliary information related to the sample.

Moreover, consider a function $\lambda_k : R^n \to R^1$ for all the units sampled and $g_k = \lambda_k(x_s)$ the value of k-th function evaluated in x_s[30].

We can divide the sample of units that participate to the survey in n groups (not necessarily disjoint) with a size given as follows[31]:

$$n_k = \sum D(g_k - g_j)(k \in s) \tag{5.57}$$

$$D(g_k - g_j) = \begin{cases} 1 & \text{if } |g_k - g_j| \leq h_k \\ 0 & \text{Otherwise} \end{cases}$$

where the value h_k is a given constant[32].

By this formulation, every group has a particular size depending upon the number of values g_j that are in the interval $[g_k \pm h_k], \forall k \in s$.

Therefore, that group has the unit k as a kernel. And its size is directly connected with the constant h_k. We can define r_q all the units sampled that respond to the survey and r all the units that are sampled and participated to the survey without knowing if they are respondent or not $(r_q \subset r)$.

We can find the absolute frequencies of sets m_k and m_{qk} as follows:

$$m_k = \sum_{j \in r} D(g_k - g_j)(k \in r) \tag{5.58}$$

$$m_{qk} = \sum_{j \in r_q} D(g_k - g_j) \qquad (k \in r_q, q = 1, \ldots, Q) \tag{5.59}$$

[29] We could have also the case where $z_k \neq x_{kq}$

[30] Assume this function exists for every x_s

[31] We call s the generic units that are sampled.

[32] The choice of h_k becomes very important in reducing the bias and the mean square error of response probabilities estimations, for instance in equation 5.61

Using equations 5.57, 5.58 and 5.59, the estimated response probabilities φ_k and ψ_{qk} can be easily found:

$$\hat{\varphi}_k = \frac{m_k}{n_k}, \forall k \in r; \qquad \hat{\psi}_{qk} = \frac{m_{qk}}{n_k}, \forall k \in r_0; \tag{5.60}$$

hence,

$$\hat{\theta}_{qk} = \hat{\varphi}_k \hat{\psi}_{qk} \frac{m_{qk}}{n_k}, (k \in r_0, q = 1, \ldots, Q); \tag{5.61}$$

The weighting adjustments are effected via nonparametric estimation of response probabilities.

5.8.3 Linear dynamic models using state-space form

It is sometimes useful to consider the linear dynamic models with another type of representation: the state-space form (SSF). Harvey (1989) represents a very useful introduction to the use of the SSF in time series analysis. An updated review of the SSF model in time series analysis is given in Durbin and Koopman (2001). Within the several advantages in using this form, we mention the followings. This form is very useful when we need to estimate an unobserved component which is time- varying. This unobserved component may contain also the regression parameters of the model, therefore this is a suitable form for estimating regression models with time- varying parameters. Furthermore, state-space form is used by the celebrated Kalman Filter (Kalman (1960)), a recursive algorithm, which has the great advantage to reduce the dimensionality problem of the estimation linked to the sample size. Kalman Filter can produce a sequential updating of the estimates and their forecasts as well the forecasts of the output variables. Here as an example we consider an application of linear dynamic models in state- space form for non-respondents units problem detected in a survey. This model is based on some estimation procedures made on previous occasions and information obtained for units responding within a fixed time. For instance, Tam (1987) introduced a linear dynamic model for analysing data from repeated surveys, discussing both the optimal predictors for finite populations and the estimation model parameters.

Let us consider a sample survey carried out at $t = 1, 2, \ldots, t_0$ on a population of N_t units $(1, 2, \ldots, N_t)$. Furthermore, assume that $y'_t = (y_{t1}, \ldots, y_{tN_t})$ is a realization of a random vector $Y'_t = (1 \times N_t)$, $X_t = (x_{t1} \ldots x_{tn_t})$ is a matrix based on prior knowledge.

At time t a sample of n_t units is surveyed from the total population (Y_t) of N_t units, so we can make a notation discriminating between the sample extracted, say Y_{ts}, and the remainder of the population, Y_{tr}; Moreover, a general l denote a column vector of N_t ones, so l_s and l_r are columns vectors comprising respectively n_t and $N_t - n_t$.

Let us consider available all the information referred to units sampled $W_t = (Y_{1s}, \ldots, Y_{ts})$ given that there is not any statistical relationship between W_{t-1} and Y_{tr}. The problem advanced by Tam consists in predicting the value of $l'Y_t$ using W_t, that reduces to the prediction of $l'_r Y_{tr}$ given all the information on $l'_s Y_{ts}$.

A generic model can be constructed as follows:

$$\begin{aligned} Y_t &= X_t \beta_t + u_t \\ \beta_t &= T\beta_{t-1} + \epsilon_t \end{aligned} \tag{5.62}$$

where β_t is an unknown vector of coefficients $(p \times 1)$. Y_t is linked with X_t by β_t unknown coefficients. In particular, β_t follows a Markovian scheme, so the 5.62 is a space state model. Assumed the joint normality of u_t and ϵ_t, the error structure is:

$$\begin{array}{lll} E(\epsilon_t \epsilon'_t) = Q_t & E(\epsilon_t \epsilon'_l) = 0 & E(\epsilon_t u'_{ts}) = L_{ts} \\ E(\epsilon_t u'_{tr}) = L_{tr} & E(\epsilon_t u'_t) = L_t & E(\epsilon_t u'_t) = 0 \end{array} \tag{5.63}$$

for $t \neq l$ $(t, l = 1, 2, \ldots, t_0)$ and T, Q_t and L_t could be considered both as known and unknown in the prediction. Tam (1987) has found the prediction of Y_{tr} and its variance, taking into consideration that β_t in the

state-space model is given by using the Kalman filter technique[33]. He also advanced further considerations for the parameters estimation of the dynamic linear model. In fact, not only β_t and Y_{tr} are unobserved, but also models parameters. A maximum likelihood estimation via nonlinear optimization procedure could be used. The estimation of the parameters and of the initial conditions of the state space models represents a challenging issue in econometrics and many approaches have been proposed in the literature. A successful approach is the Iterated Kalman Filter. Other versions of such algorithm can be found for example in Carraro and Sartore (1987). Note that Kalman-filter procedure applies not only to the linear models, but it can be used for some nonlinear and nongaussian time series models. The procedure relies upon the linearization of the nonlinear model and is called Extended Kalman Filter. We refer the interested reader to Harvey (1989).

5.9 Statistical measures and tests for rapid estimates

5.9.1 Introduction

This annex synthetically presents the most commonly used statistical measures and tests using when compiling rapid estimates and to evaluate and compare them. Obviously, the list of measures and tests presented here cannot be exhaustive since a large variety of them has been proposed in the nowcasting/forecasting literature. Nevertheless, we try here to propose the most traditionally and widely used measures and tests useful in different phases of the construction process of rapid estimates starting from the model specification until the evaluation. We assume that the reader is familiar with statistical testing theory and has a sufficient statistical background to deal with the topics discussed in this annex. For a detailed investigation on statistical testing, please refer to Hansen (2015) and Newey and McFadden (1994).

5.9.2 Stationarity tests

Augmented Dickey-Fuller (ADF) unit root test

The ADF unit root test is the generalized case of the simplest Dickey-Fuller (DF) test (Dickey and Fuller (1979)). Firstly we consider the second one by taking into analysis the following AR(1) regression model

$$y_t = \theta y_{t-1} + \epsilon_t \tag{5.64}$$

The unit root null hypothesis against the stationary alternative one corresponds to

$$H_0 : \theta = 1 \qquad \text{against} \qquad H_A : \theta < 1 \tag{5.65}$$

Alternatively, the model can be formulated as

$$\Delta y_t = (\theta - 1) y_{t-1} + \epsilon t = \phi y_{t-1} + \epsilon t, \tag{5.66}$$

where $\phi = \theta - 1 = \theta(1)$. The unit root hypothesis translates into

$$H_0 : \phi = 0 \qquad \text{against} \qquad H_A : \phi < 0 \tag{5.67}$$

The Dickey-Fuller (DF) test is simply the t-test for H_0 :

$$\hat{\tau} = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\phi}}{se(\hat{\phi})} \tag{5.68}$$

[33] For further details see in Tam (1987) pp 63-73.

where $se(\hat{\phi})$ is the standard error of the estimator $\hat{\phi}$.

The asymptotic distribution of $\hat{\tau}$ is not normal. Dickey et al. (1984) augments the basic autoregressive unit root test to accommodate general ARMA (p,q) models with unknown order (augmented Dickey-Fuller -ADF-test).

The ADF test verifies the null hypothesis that a time series y_t is $I(1)$ against the alternative that it is $I(0)$, assuming that the dynamics in the data have ARMA structure. The ADF test is based on estimating the test regression

$$\Delta y_t = \beta' D_t + \phi y_{t-1} + \sum_{j=1}^{p} \psi_y \Delta y_{t-1} + \epsilon_t \tag{5.69}$$

where D_t is a vector of deterministic terms (constant, trend, etc). The p lagged difference terms, Δy_{t-j}, are used to approximate the ARMA structure of Δy_t, and the value of p is set so that error ϵ_t is serially uncorrelated and the error term is also assumed as homoskedastic.

Under the null hypothesis, y_t is $I(1)$ which implies that $\phi = 1$. The ADF t-statistic and normalized bias statistic are based on the least squares estimates of the previous regression model and they are given by

$$ADF_t = t_{\phi=1} = \frac{\hat{\phi} - 1}{se(\hat{\phi})} \tag{5.70}$$

Phillips-Perron Test

The Phillips-Perron test (Phillips and Perron (1988)) builds on the Dickey-Fuller test. Like the augmented Dickey-Fuller test, the Phillips-Perron test addresses the issue that the process generating data for y_t might have a higher order of autocorrelation than is admitted in the test equation. Whilst the augmented Dickey-Fuller test addresses this issue by introducing lagged difference terms, Δy_{t-j} as regressors in the test equation, the Phillips-Perron test makes a non-parametric correction to the t-test statistic. The test is robust with respect to unspecified autocorrelation and heteroscedasticity in the disturbance process of the test equation. Several studies show that the Phillips-Perron test performs worse in finite samples than the augmented Dickey-Fuller test.

5.9.3 Tests on the autocorrelation function

Ljung-Box test

The Ljung-Box test (Ljung and Box (1978)) checks for the presence or absence of auto-correlation in a time series. Instead of testing the significance of the auto-correlations at each lag, it is based on all correlations up to a specified lag. It can be defined as follows. H_0: The data are independently distributed. H_1: The data are not independently distributed.

And the test statistic is:

$$Q = n(n+2) \sum_{k=1}^{k=h} \frac{\hat{\rho}_k^2}{n-k} \tag{5.71}$$

where n is the number of observations, $\hat{\rho}_k$ is the autocorrelation at lag k, and h is the number of lags being tested. Under H_0 the statistic Q follows a chi-squared distribution with h degrees of freedom.

The Ljung-Box test is commonly used in ARIMA modeling, to check that the residuals from the ARIMA model have no autocorrelation. In this case, the test must take into account the number of parameters and the Q statistics follows then a chi-squared distribution with $(h - np)$ degrees of freedom where np is the number of ARMA parameters of the model from which the residuals are derived. The Box-Pierce test statistic is a simplified version of the Ljung-Box statistic which appeared to show poor performances according to several simulation studies.

5.9.4 Seasonality tests

This section presents a set of seasonality tests that are commonly used in seasonal adjustment software. Their aim is either to a priori identify the presence of seasonality which will need further to be modelled or removed (seasonal adjustment), or to detect the presence of seasonality in the residuals of a selected model.

Friedman test (stable seasonality test)

The Friedman test is a non-parametric method for testing that samples are drawn from the same population or from populations with equal medians. In the regression equation the significance of the month (or quarter) effect is tested. The Friedman test requires no distributional assumptions. It uses the rankings of the observations.

Seasonal adjustment procedures use the Friedman test for checking the presence of seasonality. The Friedman test is called a stable seasonality test. This test uses an estimation of the Seasonal-Irregular component from which k samples are derived ($k = 12$ for monthly series and $k = 4$ for quarterly series) of size $n_1, n_2, \ldots n_k$ respectively.

Each k corresponds to a different level of seasonality. It is assumed that seasonality affects only the means of the distribution and not their variances. Assuming that each sample is derived from a random variable X_j following the normal distribution with mean m_j and standard deviation σ, the null hypothesis ($H_0 : m_1 = m_2 = \ldots = m_k$) is tested against ($H_1 : m_p = m_q$), for at least one pair (p, q).

The test uses the following decomposition of the variance:

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i,j} - \bar{x})^2 = \sum_{j=1}^{k} n_j(x_{\bullet j} - \bar{x}_{\bullet\bullet})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i,j} - \bar{x}_{\bullet j})^2, \tag{5.72}$$

where $\bar{x}_{\bullet j}$ is the average of j-th sample.

The total variance is therefore broken down into a variance of the averages due to seasonality and a residual seasonality.

The test statistic is calculated as:

$$F = \frac{\frac{\sum_{j=1}^{k} n_j(x_{\bullet j} - \bar{x}_{\bullet\bullet})^2}{k-1}}{\frac{\sum_{j=1}^{k}\sum_{i=1} nj(x_{i,j} - \bar{x}_{\bullet j})^2}{n-k}} \sim F(k-1, n-k),$$

where $k - 1$ and $n - k$ are degrees of freedom. If the null hypothesis of no stable seasonality is not rejected at the 0.10% significance level ($P_s \geq 0.001$), then the series is considered to be non-seasonal.

Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric test used for comparing samples from two or more groups. The null hypothesis states that all months (or quarters, respectively) have the same mean.

The test is calculated from an estimate of the Seasonal-Irregular component from which k samples A_j are derived ($k = 12$ for monthly series and $k = 4$ for quarterly series) of size $n_1, n_2, \ldots, n_k$ respectively.

The test is based on the statistic $W = \frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{S_j^2}{n_j} - 3(n+1)$, where S_j is the sum of the ranks of the observations from the sample A_j within the whole sample of $n = \sum_{j=1}^{k} n_j$ observations.

Under the null hypothesis the test statistic follows a chi-square distribution with $k - 1$ degrees of freedom.

Ljung-Box seasonality test

The Ljung-Box seasonality test is a direct application of the Ljung-Box test (see 5.9.3) applied to the seasonal lags. For example, for a monthly time series the test statistic is:

$$Q = n(n+2)\sum_{k=1}^{k=h}\frac{\hat{\rho}_{12k}^2}{n-k} \tag{5.73}$$

where n is the number of observations, $\hat{\rho}_{12k}$ is the autocorrelation at lag $12k$, and h is the number of lags being tested. Under H_0 the statistic Q follows a chi-squared distribution with h degrees of freedom.

Test on seasonal dummies

This test is usually based on a regression on seasonal dummies with ARIMA errors and checks the presence of deterministic and stable seasonality (Lytras et al. (2007)).

The model for a monthly series has the following general expression:

$$(1-B)^d(Y_t - \beta_1 M_{1,t} - ldots - \beta_{11}M_{1,t} - \gamma X_t) = \mu + \theta(B)\alpha_t \tag{5.74}$$

Where the regressors $M_{j,t}$ are:

$$M_{j,t} = \begin{cases} 1 & \text{in month} \quad j = 1,\ldots,11 \\ -1 & \text{in December} \\ 0 & \text{Otherwise} \end{cases} \tag{5.75}$$

Usually, the errors are supposed to follow a random walk model $(0,1,1)(0,0,0)$.

It is possible to use the individual t-statistics to assess whether seasonality for a given month is significant, or a chi-squared test statistic to check if the parameters are all equal to 0. The chi-squared test statistic is given by $\chi^2 = \hat{\beta}'[Var(\hat{\beta}')^{-1}]\hat{\beta}$and follows under the null hypothesis a χ^2 distribution with 11 degrees of freedom for a monthly series. Since the variance of the parameters, $Var(\hat{\beta})$, is computed from the estimated variance of the residuals α_t, it may be very different from the actual variance for short time series. The test is therefore corrected using the following F^M statistic:

$$F^M = \frac{\hat{\chi}^2}{11} \times \frac{n-d-k}{n-d}, \tag{5.76}$$

where n is the number of observations, d is the differencing order and k is the total number of regressors in the regARIMA model, including the 11 seasonal dummies $M_{j,t}$ and the intercept. Under the null hypothesis (no seasonality), this statistic follows a $F_{11,n-d-k}$ distribution.

Moving seasonality test

The evolutive seasonality test is based on a two-way analysis of variance model. The model uses the values from complete years only. For the Seasonal-Irregular component it uses one of the following models, depending on the type of the decomposition:

- Multiplicative: $|SI_{ij} - 1| = X_{ij} = b_i + m_j + e_{ij}$,
- Additive: $|SI_{ij}| = X_{ij} = b_i + m_j + e_{ij}$

where:

- m_j refers to the monthly or quarterly effect for j-th period, $j = (1,\ldots,k)$ where $k = 12$ for monthly series and $k = 4$ for quarterly series;

- b_j refers to the annual effect $i(i = 1, \ldots, N)$ where N is the number of complete years;
- e_{ij} represents the residual effect.

The test is based on the decomposition $S^2 = S_A^2 + S_B^2 + S_R^2$ where:

$S^2 = \sum_{j=1} k \sum_{i=1} N(\bar{X}_{ij} - \bar{X}_{\bullet\bullet})^2$ — the total sum of squares,

$S_A^2 = N \sum_{j=1} k(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet})^2$ — the inter-month (inter-quarter, respectively) sum of squares, which mainly measures the magnitude of the seasonality,

$S_B^2 = k \sum_{i=1} N(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^2$ — the inter-year sum of squares, which mainly measures the year-to-year movement of seasonality,

$S_R^2 = \sum_{j=1} k \sum_{i=1} N(\bar{X}_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet})^2$ — the residual sum of squares.

The null hypothesis H_0 is that $b_1 = b_2 = \ldots = b_N$ which means that there is no change in seasonality over the years.

This hypothesis is evaluated by the following test statistics:

$$F^M = \frac{\frac{S_b^2}{(n-1)}}{\frac{S_R^2}{(n-1)(k-1)}} \tag{5.77}$$

which is assumed to follow the F-distribution with $k-1$ and $n-k$ degrees of freedom.

HEGY test

The HEGY test (Hylleberg et al. (1990)) for seasonal integration is conducted by estimating the following regression in the case of quarterly data:

$$\Delta^4 Y_t = \alpha + \beta t + \sum_{j=2}^{4} b_j Q_{jt} + \sum_{i=1}^{4} \pi_i W_{it-1} + \sum_{l=1}^{4} \gamma_l \Delta^4 Y_{t-l} + a_t \tag{5.78}$$

Where Q_{jt} is a seasonal dummy and the W_{it} are given by:

$$\begin{aligned} W_{1t} &= (1+L)(1+L^2)Y_t \\ W_{2t} &= -(1-L)(1+L^2)Y_t \\ W_{3t} &= -(1-L)(1+L)Y_t \\ W_{4t} &= -L(1-L)(1+L)Y_t = W_{3t-1} \end{aligned}$$

After OLS estimation, tests are conducted for $\Pi_1 = 0$, for $\Pi_2 = 0$ and a joint test of the hypothesis $\Pi_3 = \Pi_4 = 0$. The HEGY test is a joint test for zero frequency and seasonal unit roots. If none of the Π_i are equal to zero, then the series is stationary (both at seasonal and nonseasonal frequencies). The HEGY test has been extended for monthly series by Franses (1991) and Beaulieu and Miron (1993).

Canova-Hansen test

The test developed by Canova and Hansen (1995) takes as the null hypothesis that the seasonal pattern is deterministic. From the following model:

$$\begin{aligned} y_t &= \alpha y_{t-1} + \sum_{i=1}^{S-1} D_{i,t}\beta_{i,t} + \epsilon_t \\ \beta_{i,t} &= \beta_{i,t-1} + u_t \end{aligned} \tag{5.79}$$

they test that the variance of the u_t is equal to 0. The null hypothesis in the Canova-Hansen test is therefore rejected in case seasonality of a series is not constant. The Canova-Hansen test should not be rejected after seasonal adjustment or when the series has no seasonal pattern, which also implies constant seasonality. Canova-Hansen suggest a Lagrange Multiplier test statistic whose distribution is known as von Mises distribution. The test is rejected for the large values of L-statistics. Canova and Hansen use the assumption that both the process under investigation and the explanatory variables in the null regression do not contain any non-stationary behavior at the zero frequency.

5.9.5 Information criteria (AIC, BIC)

Information criteria are very useful statistical measures in many aspects of the modelling and testing. Relevant applications are in selecting the set of regressors or as criteria for model selection. For example, they are used to choose between several ARIMA models for a time series or in the ADF test in order to choose lag-length of the differenced variables which enter as regressors in the model. These criteria are mainly based on the variance of the residuals of the model: the smaller the variance, the better the fit and the model. They also introduce a term to take into account the parsimony of the model, and the number of observations.

The two main measures are the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC).

$$\begin{aligned} AIC(p,q) &= Log\hat{\sigma}^2 + \frac{2(p+q)}{n} \\ BIC(p,q) &= Log\hat{\sigma}^2 + (p+q)\frac{Logn}{n} \end{aligned}$$

As it is easy to see from the definitions, the smaller the criterion, the better the model. Model with too many parameters regarding to the number of observations are penalized. Many other criteria have been proposed: the corrected AIC (AICC), the corrected BIC (BICC), Hannan-Quin criterion etc.

5.9.6 Granger causality test

Consider two time series x_t and y_t; A researcher may want to test the hypothesis that x_t causes y_t. Therefore, a causality test based on the null hypothesis of no causality between the two time series should be set up. This approach is very useful in time series analysis and in particular, in forecasting, when one economic variable can help to forecast another one. Granger (1969) was the first arguing that, considering two time series, their underlying generation process can be an infinite-order autoregressive process.

Moreover, he was interested in testing if any lagged information on time series y_t provides any statistically significant information about the other series x_t. If not, then "y_t does not Granger-cause x_t." Assume that the autoregressive generating process has a k order, the Granger test is based on OLS estimation of this equation:

$$x_t = \sum_{i=1}^{k} \alpha_i x_{t-i} \sum_{i=1}^{k} \beta_i y_{t-j} + u_t \tag{5.80}$$

The null hypothesis is based on: $\beta_1 = \beta_2 = \ldots = \beta_k$. So, if the null hypothesis is verified, y_t does not cause x_t.

This null hypothesis is tested by means of an F-statistic, taking into consideration the sum square residuals of the unrestricted model (URSS) in equation 5.80 and the sum square residual of the restricted one (RRSS)[34] as follows:

$$\frac{(RRSS - URSS)/k}{URSS/n - 2k - 1} \sim F_{(k,n-2k-1)} \tag{5.81}$$

Where k is the number of restrictions and $n - 2k - 1$ are the degrees of freedom[35]. If the value obtained from this statistic is greater that the specific critical value, the null hypothesis is rejected and Granger causality has been found.

5.9.7 Evaluating the quality of a regression model

When using regression models for building up a rapid estimate, several aspects should be considered when assessing the quality of the model:

- The goodness of fit, that can be assessed using the R-squared and associated statistics, the F-statistic to verify the global significance of the regressors and the Student statistics to assess the significance of each regressor;
- The absence of misspecification of the model. It is for example important to pay attention to the structure of the residual - absence of auto-correlation (Durbin and Watson (1950),Durbin and Watson (1951)), absence of heteroskedasticity (Breusch and Pagan (1979)) - and the possible collinearity between regressors (Besley et al. (1980) statistics).

All these tests and statistics can be used for other kind of models.

The *coefficient of determination* (R^2) is the proportion of a sample variance of a response variable that is "explained" by the predictor (explanatory) variables when a linear regression is done.

The formula for R^2 is:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \tag{5.82}$$

where ESS is the explained sum of squares, RSS is the residual sum of squares, and TSS is the total sum of squares.

R^2 always increases when a new variable is added to a model, unless the new variable is perfectly multi-collinear with the original variables[36].

5.9.8 Johansen cointegration test

The cointegration problem is analysed with the Johansen cointegration test.

Consider a vector time series process y_t composed by some components $y_{i,t}$, for $i = 1, 2, \ldots, k$, that are unit root processes, there is cointegration if certain linear combinations of $y_{i,t}$ are stationary.

Johansen (1988) has considered a VAR model of order p:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + B x_t + \epsilon_t, \tag{5.83}$$

[34] $x_t = c \sum_{i=1}^{k} \gamma_i x_{t-i} + \epsilon_t$

[35] n is the number of observations.

[36] The adjusted R^2 (i.e. $\bar{R}^2$) penalizes the R^2 for the addition of regressors which do not contribute substantially to the explanatory power of the model. It is computed as: $\bar{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-k}$

where y_t is a k-vector where, x_t is a vector of deterministic variables and $\epsilon_t \sim nid(0, \sum)$ (i.e is normal, independent and identically-distributed process). Another way to write the model 5.83 is:

$$\Delta y_t = \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + B x_t + \epsilon_t, \tag{5.84}$$

where $\Pi = \sum_{i=1}^{p} A_i - I$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j$.

Therefore, if the coefficient matrix Π has reduced rank $r(k > r)$, there exists $k \times r$ matrices α and β with rank r (matrices of full rank). The reduced rank of matrix Π is considered as a cointegration hypothesis (where r is the order of cointegration) and, if satisfied, implies $\Pi = \alpha\beta'$ and $\beta' y_t$ is I(0).

Johansen has proposed two tests for the number of cointegrated vectors.

The first one (Lambda-max test) is based on the log-likelihood ratio test $\ln(\mathsf{L}_{max}(r_0)/\mathsf{L}_{max}(r_0+1))$ and more precisely $-n\ln(1-\hat{\lambda}_{r_0+1}) = n\hat{\lambda}_{r_0+1}$ is computed sequentially[37]. The value r_0 represents the number of cointegration vectors in the null hypothesis to be tested: $H_0(r_0) : r = r_0$ against $H_1(r_0) : r_0 = r_0 + 1$. If the null hypothesis is rejected there could be $r_0 + 1$ cointegrating vectors. This test is performed sequentially till the null hypothesis is accepted.

The second one (Trace test) is based on the log-likelihood ratio $\ln(\mathsf{L}_{max}(r_0)/\mathsf{L}_{max}(q))$ and $-n\sum_{i=r_0}^{q}\ln(1-\hat{\lambda}_i)$ is computed sequentially. In this case the null hypothesis is as follows: $H_0(r_0) : r = r_0$ against $H_1(r_0) : r_0 = r > r_0$. If the null hypothesis is rejected, there are more than r_0 cointegrating vectors. For instance if we set $r_0 = 0$, the alternative hypothesis becomes $r > 0$. In this particular case, if the null hypothesis is rejected, at least one cointegrating vector is detected. Successively, we set $r_0 = 1$. If the null hypothesis is accepted, only one cointegrating vector is found, otherwise, we should proceed sequentially until the null hypothesis is not rejected (for instance setting $r_0 = 2$ and so on).

5.9.9 Chow inequality test

Sometimes the application of the same regression model or the same coefficient to a particular sample of data may not be the right choice. The structure of the model or the coefficients might change within the sample.

Therefore, when structural change occurs, the same coefficients or the same regression model cannot be applied for the whole period. The Chow test (see Chow (1960)) was introduced with the particular aim of detecting possible structural change within time series (change in parameters coefficient).

Consider a sample (y and x) of data where a researcher suspects that a structural change has occurred; this sample should be divided into two sub-samples (y_1 and y_2, x_1 and x_2).

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \tag{5.85}$$

The unrestricted least square estimation is:

$$\hat{\beta} = (x'x)^{-1}x'y = \begin{pmatrix} x_1'x_1 & 0 \\ 0 & x_2'x_2 \end{pmatrix}^{-1} \begin{pmatrix} x_1'y_1 \\ x_2'y_2 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \tag{5.86}$$

so, two residual sums of square would result from the two regressions. Therefore the total sum of square residual should be the sum of the residual of the two regressions. The researcher should set up a model

[37] We refer to $\hat{\lambda}$ as the estimated eigenvalue of matrix Π.

where the two coefficients for the different periods are the same. In other words, he should construct a restricted model where $\beta_1 = \beta_2$.

The sum of square residual of the restricted model (RRSS)[38] should be compared with the sum of square residual of the previous unrestricted model (URSS)[39] where the possible structural change has been taken into consideration:

$$\frac{(RRSS - URSS)/q}{URSS/n-k} \sim F_{(q,n-k)} \tag{5.87}$$

where n is the sample size. The statistical test is performed by an F-test with q restrictions (in the previous example $q = 1$) and $n - k$ degrees of freedom (in the previous example $k = 2$). If a structural change has occurred, the hypothesis that $\beta_1 = \beta_2$ would be rejected and the inequality between the parameters would be proven.

5.9.10 Forecast accuracy measures and tests

Suppose that our reference variable y_t has been modelled over the period $t = 1 \ldots T$ following one or several modelling strategies. The main approach for evaluating the forecast accuracy of a model is to construct sequentially out-of-sample forecasts and to compare them with the realized values of the reference variable. Here we are restricting our analysis to the case of one-step-ahead forecast which is the most natural way to build up rapid estimates.

Consider the realized value of our reference variable y_{T+k+1}, where $k = 0, \ldots, r-1$ and $r > T$, and the corresponding one-step-ahead forecast sequentially computed $\hat{y}_{T+k+1}$, we define the forecasting error as: $e_{T+k+1} = y_{T+k+1} - \hat{y}_{T+k+1} = e_j = y_j - \hat{y}_j$, where $j = T + k + 1$ varies with k from $T + 1$ to $T + r$.

Forecast errors are the basis of several measures and tests for evaluating the accuracy of the results or to compare alternative forecasts.

Measuring forecast accuracy

Below we are presenting the most commonly used measures to evaluate a forecast or to compare forecasts of the same reference variable, generated by different models. Unfortunately, those measures provide just indicative information but they do not tell us if a given forecast is good or which forecast among a set of competing ones is outperforming in a significant way the others.

Mean forecasting error

The mean forecasting error (MFE) is the average of forecasting errors computed over a given horizon. It provides an indication of the presence of bias in our forecasts. When the MFE is zero or near to zero, there is no bias or a very small one, while higher values of the MFE provide a signal of an important bias in the computed forecasts.

Mean percentage forecasting error

The mean percentage forecasting error (MPFE) is the mean of percentage forecast errors over a given time horizon.

The mean percentage forecasting error is given by:

$$MPFE = \frac{100}{r} \sum_{j=T+1}^{T+r} \frac{y_j - \hat{y}_j}{y_j} \tag{5.88}$$

[38] It is considered the particular restriction (in the example is $\beta_1 = \beta_2$)
[39] The model considered in the equation 5.86

Root-mean-square forecasting error

The root-mean-square forecasting error (RMSFE) is probably the most frequently used measure of forecasting accuracy. The RMSFE represents the sample standard deviation of the differences between predicted values and observed values.

The RMSFE is given by:

$$RMSFE = \sqrt{\frac{\sum_{j=T+1}^{T+r} (y_j - \hat{y}_j)^2}{r}} \tag{5.89}$$

Diebold-Mariano test

Let suppose that for a given target variable y we derived two competing forecasts stemming from different models $\hat{y}_1$ and $\hat{y}_2$. By recursively computing one-step-ahead forecasts over the period $T+1, \ldots, T+r$, we obtain two series of forecasts $\hat{y}_{1,j}$, and $\hat{y}_{2,j}$, $j = T+1, \ldots, T+r$. The associated forecasting errors are respectively $e_{1,j}$, and $e_{2,j}$. When comparing the two alternative forecasts, we are essentially interested in knowing if they are significantly different or not. The Diebold-Mariano test (Diebold and Mariano (1995)) also known as equal accuracy test provides an answer to this.

The Diebold-Mariano test is given by:

$$S = \frac{\frac{1}{r} \sum_{j=T+1}^{T+r} \{g(e_{1,j}) - g(e_{2,j})\}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{r}}} \tag{5.90}$$

with $g(e_{i,j})$, $i = 1, 2$ denoting the loss from forecast error $g(e_{i,j})$ evolving from prediction model i. The null hypothesis tested is $H_0 : Eg(e_{1t}) = Eg(e_{2t})$. Under H_0, S is asymptotically standard normal distributed, $S \approx N(0, 1)$.

Harvey et al. (1997) derived a finite sample distribution of the Diebold-Mariano test. This test is very widely used and sometimes abused. A critical analysis of the use of the Diebold-Mariano test has been provided by Diebold (2013).

5.9.11 Encompassing tests for discriminating two competitive models in terms of predictive ability

The encompassing test is a useful tool to find which of several models could be considered the best in terms of predictive ability. Granger and Ramanathan (1984) suggested a regression-based method, by which the best forecasting model should be chosen. A researcher may consider two competing forecasts $\hat{y}_{1,j}$ and $\hat{y}_{2,j}$, $j = T+1, \ldots, T+r$ for y_j and construct a model as follows:

$$y_j = \beta_0 + \beta_1 \hat{y}_{1,j} + \beta_2 \hat{y}_{2,j} + \epsilon_j \tag{5.91}$$

In this model, $\hat{y}_{1,j}$ or $\hat{y}_{2,j}$ are responsible (by the driving effect of β) for the forecast of y in j. A discrimination between the two forecasts can be carried out by an encompassing test. A simple OLS regression-method is used for estimating weights β_0, β_1, β_2.

Hendry and Chong (1986) define the null hypothesis that the forecast for model 1 encompasses the forecast for model 2 if β_1 and β_2 are respectively 1 and 0.

Therefore, the test for encompassing could be expressed as follows:

$$\hat{u}_{1,j} = \beta_0 + \beta_2 \hat{y}_{2,j} + \epsilon_j \tag{5.92}$$

Where, $\hat{u}_{1,j} = y_j - \hat{y}_{1,j}$ and OLS estimation is performed. By t-statistic associated with β_2 a researcher could test the null hypothesis that model 1 encompasses model 2. For more details on the use of the encompassing test and its links to forecast combination, please see Costantini and Pappalardo (2008), Harvey et al. (1998) and Fang (2003).

Bibliography

Akaike, H. (1981), "Likelihood of a model and information criteria", Journal of Econometrics 16, 1, 3-14.

Al-Osh, M. (1989),"A dynamic linear model approach for disaggregating time series data", Journal of Forecasting, 8, pp. 65-96.

Andren, T. (2007), "Econometrics", textbook, BookBoon, 155p.

Baffigi, A., Golinelli, R., Parigi, G. (2004), "Bridge Models to Forecast the Euro Area GDP", International Journal of Forecasting, forthcoming.

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.

Bardsen, G. (1989),"The Estimation of Long-run Coefficients from Error- correction Models" Oxford Bulletin of Economics and Statistics, 51, 345-350.

Barigozzi, M., Lippi, M., Luciani, M., (2014), "Dynamic Factor Models, Cointegration, and Error Correction Mechanisms", ECARES working paper, 2014-14.

Bartlett, M.S. (1948), "Smoothing periodograms for time series with continuous spectra," Nature 161, 686-687. (reprinted in [Kesler 1986]).

Bassie, V.L. (1958), Economic Forecasting, McGraw-Hill, New York.

Battaglia, F., Fenga L. (2003): "Forecasting composite indicators with anticipated information: an application to the industrial production index", Journal of the Royal Statistical Society, Applied Statistics, 52, 3, 279-290.

Beaulieu, J.J., Miron, J.A. (1993) "Seasonal Unit Roots In Aggregate U.S. Data," Journal of Econometrics, v55(1/2), 305-328

Bell, W.R., Hillmer, S.C. (1990), "The time series approach to estimation for repeated surveys", Survey Methodology, 16, 195- 215

Besley, D.A., Kuh, K., Welsch, R.E. (1980), "Regression diagnostics: Identifying influential data and sources of collinearity", John Wiley and Sons, New York

Beveridge, S., Nelson, C. (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle", Journal of Monetary Economics, 7, 151-174.

Blake, A., Kapetanios, G., Weale, M. (2000), "Industrial Production Nowcasting". Report submitted to EUROSTAT by National Institute of Economic and Social Research.

Bodo, G., Signorini, L. (1987) "Short-term forecasting of the industrial production index", International Journal of Forecasting, 10, 285-299.

Boot, J.C.G., Feibes, W., Lisman, J.H.C. (1967), Further Methods of Derivation of Quarterly Figures from Annual Data, Applied Statistics 16 (1).

Bournay, J., Laroque, G. (1979), "Réflexions sur la méthode d'élaboration des comptes trimestriels", Annales del'INSEE, 36.

Bovi, M., Lupi, C., Pappalardo, C. (2000), "Predicting GDP Components Using ISAE Bridge Equations Econometric Forecasting Model (BEEF)", Documento di Lavoro ISAE, n. 13. Rome.

Box, G.E.P., Jenkins, G.M. (1976), Time Series Analysis Forecasting and Control,original edition (1970), Holden-Day, San Francisco.

Breusch, T.S., Pagan A.R. (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation", Econometrica, 47, 1287–1294.

Bruno, G., Lupi C. (2004), "Forecasting industrial production and the early detection of turning points", Empirical Economics, Springer, 29, 3, 647-671.

Camba-Mendez, G., Kapetanios, G., Smith, R., Weale, M. (2001), "An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries", Econometric Journal.

Canova, F., Ciccarelli, M. (2013), "Panel Vector Autoregressive Models: A Survey", Advances in Econometrics, volume 31.

Canova, F., Hansen B.E. (1995), "Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability", Journal of Business and Economic Statistics, 13, 237-252.

Carraro, C., Sartore, D. (1987), Square Root Iterative Kalman Filter: Theory and Applications to Regression Models, Annales d'economie et de statistique, 6/7, INSEE, Paris, 1987.

Chudik, A., Pesaran, M.H. (2014), "Theory and Practice of GVAR Modeling", Working Paper No. 180, Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute.

Chen, J., Shao, J. (2000), "Nearest Neighbour Imputation for survey data", Journal of Official Statistics, 16, 113-131.

Chow, G.C (1960),"Tests of Equality Between Sets of Coefficients in Two Linear Regressions", Econometrica, 28, 591-605.

Chow, G., Lin, A.L. (1971),"Best Linear Unbiased Interpolation, Distribution and Extrapolation of time series by Related Series", The Review of Economics and Statistics, 53(4), 372-375.

Clemen, R.T. (1989), "Combining forecasts: A review and annotated bibliography", International Journal of Forecasting 5 (1989), 559-583.

Cochrane, J.H. (2005), "Time Series for Macroeconomics and Finance", Graduate School of Business, University of Chicago.

Costantini, M., Pappalardo, C. (2008), "Combination of Forecast Methods Using Encompassing Tests. An Algorithm-Based Procedure" Economics Series 228, Institute for Advanced Studies.

Crane, D.B., Crotty, J.R. (1967), "A two-stage forecasting model: Exponential smoothing and multiple regression", Management Science, 13, B501-B507.

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 (333).

Dickey, D.A., Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association 74, 427-431.

Dickey, D.A., Hasza, D.P., Fuller, W.A. (1984), "Testing for Unit Roots in Seasonal Time series", Journal of American statistical Association, 79, 355-367.

Diebold, F.X., Mariano R.S. (1995), "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, 13, 253-263.

Diebold, F.X. (2013), "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests", University of Pennsylvania.

Di Fonzo T. (1987), "La stima indiretta di serie economiche trimestrali", Padova, Cleup.

Diron, (2008), "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data," Journal of Forecasting, John Wiley and Sons, Ltd., vol. 27(5), pages 371-390.

Durbin, J., Koopman, S.J. (2001), Time Series Analysis by State Space Methods, Oxford Univerity Press, New York.

Durbin, J., Watson, G.S. (1950), "Testing for Serial Correlation in Least Squares Regression, I". Biometrika 37 (3-4): 409-428.

Durbin, J., Watson, G.S. (1951), "Testing for Serial Correlation in Least Squares Regression, II". Biometrika 38 (1-2): 159-179.

Eltinge, J.E., Yansaneh I.S. (1997), "Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. consumer expenditure survey", Survey Methodology, 23, 1, 33-40.

Engle, R., Watson, M. (1981), "A One Factor Multivariate Time Series Model of Metropolitan Wage Rates", Journal of the American Statistical Association 76, 774-781.

Engle, R., Granger, C.W.J. (1987) "Co-integration and error correction: representation, estimation and testing", Econometrica, 55, 251-276.

Entfor, H. (1992), "Random Walk with Drift, Simultaneous Errors, and Small Sample: Simulating the Bird's Eye View", Institut National de la Statistique et des Etudes Economiques.

Ericsson N.R., Reisman E.L. (2012), "Evaluating a global vector autoregression for forecasting", International Finance Discussion Papers 1056, Board of Governors of the Federal Reserve System (U.S.).

Eurostat, Handbook on Quarterly National Accounts (1999).

Fang, Y. (2003), "Forecasting combination and encompassing tests", International Journal of Forecasting 19 (2003), 87-94.

Fernandez, R.B. (1981), "A methodological Note on the Estimation of Time Series", The Review of Economics and Statistics 63 (3).

FLASH software, Reference Manual, 2002

Foroni, C., Marcellino, M., Schumacher, C. (2012): "U-MIDAS: MIDAS regressions with unrestricted lag polynomials", CEPR Discussion Papers, 8828.

Frale, C., Marcellino, M., Mazzi,G.L., Proietti, T. (2010): "Survey data as coincident or leading indicators", Journal of Forecasting, 29(1-2), 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, 439-470.

Franses, P.H. (1991), "Seasonality, Nonstationarity and the Forecasting of Monthly Time Series", International Journal of Forecasting, 7, 199-208.

Garderen, K.J., Lee K., Pesaran M. (2000), "Cross-Sectional Aggregation of Non-linear Models", Journal of Econometrics, 95, 2, 285-331.

Ghysels, E., Santa-Clara, P., Valkanov, R. (2004), "The MIDAS touch: Mixed Data Sampling regression models", mimeo.

Giannone, D., Reichlin, L., Small, D. (2008): "Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases", Journal of Monetary Economics, 55, 665-676.

Giommi, A. (1987), "Nonparametric methods for estimating individual response probabilities", Survey Methodology, si3, 2, pp. si27-si34.

Gourieroux, C., Monfort, A. (1997), Time Series and Dynamic Models, Cambridge University Press, Cambridge.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Methods and Cross-Spectral Methods", Econometrica, 34, 424-438.

Granger, C.W.J., Bates, J. (1969), "The Combination of Forecasts", Operational Research Quarterly 20, 451-468.

Granger, C.W.J., Newbold, P. (1974), "Spurious Regression in Econometrics", Journal of Econometrics, 2, 111-120.

Granger, C.W.J., Ramanathan, R. (1984), "Improved Methods of Combining Forecasting", Journal of Forecasting, 3, 197-204.

Hansen, B.E. (2015), "ECONOMETRICS", textbook, University of Wisconsin, Department of Economics.

Hatanaka, M. (1996), Time-Series-Based Econometrics, Oxford University Press.

Hendry F.H. (1995), Dynamic Econometrics, Oxford University Press.

Hendry, D.F., Chong, Y.Y. (1986),"Econometric Evaluation of Linear Macro-Economic Models", Review of Economic Studies, 53, 671-690.

Harvey, A.C. (1981), "Time Series Models", Oxford: Phillip Allan and Atlantic Highlands, NJ: Humanities Press.

Harvey, A.C., Pierse, R.G. (1984), "Estimating missing observations in economic time series", Journal of the American Statistical Association, 79, 125-131.

Harvey, A. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press.

Harvey, D.I., Leybourne, S.J., Newbold, P. (1997), "Testing the Equality of Prediction Mean Square Errors" International Journal of Forecasting, 13, 273-281.

Harvey, D.I., Leybourne, S.J., Newbold, P. (1998), "Tests for forecast encompassing." Journal of Business and Economic Statistics, 16, 254-259.

Hendry, D.F., Pagan, A.R., Sargan, J.D. (1984), "Dynamic Specification", in Z. Griliches and M.D. Intrilligator (eds.), Handbook of Econometrics II, North-Holland, Amsterdam, 1023-1100.

Hylleberg, S., Engle, R.F., Granger, C.W.J., and Yoo, B.S. (1990), "Seasonal integration and cointegration", Journal of Econometrics, 44: pages 215-238.

Johnston, J. (1984), Econometric Methods, 3rd edition, McGraw-Hill.

Johansen, S. (1988), "Statistical Analysis of Co-integration vectors", Journal of Economic Dynamics and Control, 12, 231-254.

Jones, R.H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations", Technometrics, 22, 389-395.

Kalman, R. E. (1960). A New Approach to Linear Filtering And Prediction Problems, Transactions of the ASME–Journal of Basic Engineering, serie D, 82, 35-45.

Kalton, G. (1983), "Compensating for Missing Survey Data", Michigan.

Litterman, R.B. (1983), "A random Walk, Markov Model for the Distribution of Time Series", Journal of Business and Economic Statistics 1 (2).

Little, R.J.A. (1986), "Survey Nonresponse Adjustments for Estimates of Means", International Statistical Review, 54, 2, 139-157.

Ljung, G.M., Box (1978), G.E.P., "On a Measure of a Lack of Fit in Time Series Models". Biometrika 65 (2): 297-303.

Lütkepohl, H. (2006), "New Introduction to Multiple Time Series Analysis", Springer-Verlag, New York.

Lytras, D.P., Feldpausch, R.M., Bell, W.R. (2007), Determining Seasonality: A Comparison of Diagnostics From X-12-ARIMA, Papers presented at the ICES-III, June 18-21, 2007, Montreal, Quebec, Canada

Maravall, A. (1983), "An Application of Nonlinear Time Series Forecasting", Journal of Business and Economic Statistics, 1, 66-74.

Mariano, R., Murasawa, Y. (2003): "A new coincident index of business cycles based on monthly and quarterly series", Journal of Applied Econometrics, 18(4), 427-443.

Mariano, R., Murasawa, Y. (2010): "A coincident index, common factors, and monthly real GDP", Oxford Bulletin of Economics and Statistics, 72(1), 27-46.

di Mauro, F., Pesaran M.H. (2013), "The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis", Oxford University Press.

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., Savio, G. (2005), "Theory and applications of univariate and multivariate models for temporal disaggregation". Working paper. European Commission. 25p.

Mitchell, J., Smith, R., Weale, M., Wright S., Salazar E.L. (2005), "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth". Royal Economic Society, vol. 115(501), pages F108-F129, 02.

Moauro, F., Savio, G. (2005), "Temporal disaggregation using multivariate structural time series models." The Econometrics Journal 8, 214- 34.

Nelson, C.R., Plosser, C.I. (1981), "Trends and random walks in macroeconomic time series: Some evidence and imphcations", Journal of Monetary Economics 10, 139-162.

Newey, W. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R. and McFadden, D., Eds., Handbook of Econometrics, Elsevier, Berlin, 2111-2245.

Niyonsenga, T. (1994), "Nonparametric Estimation of Response Probabilities in Sampling Theory", Survey Methodology, 20, 2, pp. 177-184.

Oh, H.L., Scheuren, F.S. (1983), "Weighting Adjustments for Unit Nonresponse", in: Incomplete Data in Sample Surveys, Vol. 2, Theory and Bibliographies, Madow, W.G., Olkin, I., Rubin, D.B. eds., Academic Press, New York. pp. 143-184.

Parzen, E., (1957), "On the Consistent Estimates of the Spectrum of a Stationary Time Series," Annals of Mathematical Statistics 28, 329-48.

Pesaran, M.H., Schuermann, T., Weiner, S.M.(2004), Modelling regional interdependencies using a global error-correcting macroeconometric model. Journal of Business and Economics Statistics 22,129-162.

Pesaran M.H., Schuermann T., Smith L.V. (2008), "Forecasting Economic and Financial Variables with Global VARs", CESifo Working Paper Series 2263, CESifo Group Munich.

Phillips, P.C.B. (1986), "Understanding Spurious Regression in Econometrics", Journal of Econometrics,33, 311-340.

Phillips, P.C.B., Perron, P. (1988), "Testing for a Unit Root in Time Series Regression". Biometrika 75 (2): 335-346.

Plosser, C.I., Schwert, G.W. (1977), "Estimation of a Noninvertible Moving Average Process: The Case of Overdifferencing," Journal of Econometrics, 6, 199-224.

Poissonnier, A., Cicconi, C., Buono, D. and others, (2013 edition), Chapter 5 of Handbook on quarterly National Accounts, Eurostat

Rizzo, L., Kalton, G., Brick, J.M. (1996), "A Comparison of Some Weighting Adjustment Methods for Panel Nonresponse". Survey Methodology, 22(1), 43-53.

Salazar, E., Smith, R., Weale, M. (1997), "Interpolation by means of a regression model: estimation and Monte-Carlo analysis", National Institute of Economic and Social Research Discussion Paper No. 126.

Santos Silva, J.M.C., Cardoso, F. (2001) "The Chow-Lin method using dynamic models", Economic Modelling, vol. 18, p. 269-280

Sims, C. (1980), "Macroeconomics and Reality", Econometrica, 48, 1-48.

Schorfheide, F., Song, D. (2011): "Real-time forecasting with a mixed frequency VAR", mimeo.

Statistics Sweden (2000), An Activity Index for the Swedish economy.

Stock, J.H. (1987), "Asymptotic properties of a least squares estimation of co-integrating vectors, Econometrica, 55, 1035-1056

Stock, J., Watson, M. (1989), "New Indexes" NBER Macroeconomics Annual 4, 351-394.

Stock, J., Watson, M. (1998), "Diffusion Indexes" NBER Working Paper Series 6702.

Stock, J., Watson, M. (2002), "Macroeconomic Forecasting Using Diffusion Indexes" Journal of Business and Economics Statistics v20(2, Apr), 147-162.

Stock, J., Watson, M. (2010A), "Dynamic Factor Models", in: Oxford Handbook of Economic Forecasting, Michael P. Clements and David F. Hendry (eds), Oxford University Press.

Stock, J., Watson, M. (2010B), "Introduction to Econometrics", Addison- Wesley; 3rd edition (December 13, 2010), 840p.

Stram D.O., Wei W.W.S. (1986), "A methodological Note on the Disaggregation of Time Series Totals", Journal of Time Series Analysis 7 (4).

Tam, S.M. (1987), "Analysis of repeated surveys using a dynamic linear model". International Statistic Review, 55, 63-73.

Tukey, J.W., Blackman, R.B., (1959), "The Measurement of Power Spectra", Dover Publications Incorporated.

Van der Ploeg F. (1985), "Econometrics and Inconsistencies in the National Accounts", Econometric Modelling 2(1), pp.8-16.

Wei W.W.S., Stram D.O. (1990), "Disaggregation of Time Series Models", Journal of Royal Statistical Society, 52 (3).

Zarnowitz, V. (1967), "An appraisal of short-term economic forecasts" National Bureau of Economic Research, New York.

6 A Review of Parametric and Non-parametric Techniques for Variable Selection

Authors

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

6.1 Introduction

The quest for the best model, the one that will give good and reliable predictions, has been the object of decades of research and hundreds of papers. One of the key aspects in the construction of an econometric model, in particular in the context of rapid estimation and leading indicators, is of course the selection of the explanatory variables and this is still a major open research topic. Nowadays, with the increasing dimension of datasets due to the larger data availability, this problem has become of crucial importance.

As there is a large number of possible indicators when lags are also considered the number of candidate models can be measured in terms of millions. Choosing between them is subject to high search costs, so the modeller cannot be confident that the best representation of the true data generating process has been selected.

The problem is further complicated by a lack of degrees of freedom as the number of indicators approaches or exceeds the number of time observations, and multicollinearity due to a high correlation between rival variables. Apart from the 'traditional approach' based on trials and errors starting from a pre-specified model chosen for its economic relevance, most of the various automatic approaches fall into two broad categories: data reduction and reduced regression.

Data reduction is based on the notion that a given pool of indicators can be adequately represented by a subset of the data. Cluster analysis suggests that indicators can be grouped together according to a certain measure of similarity, and hence smaller numbers of candidate variables can be selected from each cluster. Factor analysis implies 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 principal components or dynamic factor models. It is important to note that in these data reduction techniques, the target variable does not directly participate to the variable grouping process (*unsupervised learning*).

With reduced regression the target variable is directly used to select the adequate explanatory variables (*supervised learning*), taking also into account the likely collinear structure of the variables. Among the most popular methods are the Forward Selection, Backward Elimination, Forward Stepwise Regression, Least Angle Regression (LARS), Least Absolute Shrinkage and Selection Operator Regression (LASSO), Elastic net etc., and of course the General-to-specific modelling which starts from a general and over-specified statistical model that captures the essential characteristics of the underlying data set, and then uses standard testing procedures to reduce its complexity by eliminating insignificant variables to form a more parsimonious model of the data. Some 'hybrid methods' make the link between supervised and unsupervised learning: Reduced-Rank Regression, Partial Least Squares, etc.

The variable selection problem has also been studied in the Bayesian context and specific methods originate from this literature: Stochastic search variable selection (SSVS), Reversible jump MCMC, Composite model space (CMS) etc.

Very good papers reviewing and comparing these methods already exist, from Hocking (1976) to Ng (2013). This chapter provides a careful and critical review of the available econometric and statistical methods to handle variable selection in the presence of large dimensional datasets. Section 6.2 presents the statistical framework and the notations used in the chapter. Section 6.3 discusses reduced regressions and section 6.4 deals with data reduction methods.

6.2 Statistical Framework and Notations

In this chapter, we will use linear regression models as the main statistical framework. We have a target variable Y, the Euro-area GDP for example, we want to predict using a set of N continuous or discrete explanatory variables $(X_1, X_2, \ldots, X_N)$, such as the industrial production indices, the turnover indices, the business tendency surveys indicators, an outlier, a trading-day effect etc.

The linear regression model has the form:

$$f(X) = \beta_0 + \sum_{i=1}^{i=N} \beta_i X_i$$

Linear regression assumes that the regression function $E(Y|X)$ is linear in the inputs $X = (X_1, X_2, \ldots, X_N)$, or that the linear model is a reasonable approximation. Linear models are simple and often provide an adequate and interpretable description of how the inputs X affect the response function Y. The β_i are unknown parameters that will be estimated from a set of training data $(x_1, y_1), (x_2, y_2), \ldots, (x_T, y_T)$ where y_t is the value of the target variable Y at date t and $x_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})$ the vector of the covariate values at date t. The most popular estimation method is least squares, in which the unknown parameters are chosen to minimize the residual sum of squares

$$RSS(\beta) = \sum_{t=1}^{t=T} [y_t - f(x_t)]^2 = \sum_{t=1}^{t=T} \left[y_i - \left(\beta_0 + \sum_{i=1}^{i=N} \beta_i X_{it} \right) \right]^2$$

In matrix notation, this minimization problem and its well-known solution can be expressed as:

$$\min_{\beta} [y - X\beta]^2 \; with \; Y = X\beta + \varepsilon$$

$$\hat{\beta} = (X^{'}X)^{-1}X^{'}y$$

$$\hat{y} = X\hat{\beta} = X(X^{'}X)^{-1}X^{'}y$$

It is also important to note that in a linear regression model, you must have less explanatory variables than observations: N must be smaller than T. This is usually not a problem for survey data, where you often have thousands of observations, but it can be more problematic for economic forecasting models where you might only have ten years of observations for hundred of potential regressors.

The linear model relies on the independency of the explanatory variables, hypothesis scarcely met when data do not come from an experimental design. In particular, in the case of economic forecasting, regressors usually present collinearity and least square estimates, if they have virtually no bias, might have a large variance which conducts to not so accurate forecasts. The forecast quality could sometimes be improved by shrinking or setting some coefficients to zero. In this case, we can accept a small bias to improve the quality of the model by reducing the variance of the predicted values. This strategy conducts to focus on a smaller set of regressors which exhibit the strongest effects, and on a model that captures the 'big picture' and can be easily interpreted and commented. As we will see, several methods have been proposed to correct for these problems.

At the end of the day, each variable selection method will propose one or several regression model and you need a statistical criterion to measure and compare the quality of these models. The R-square is not a very good criterion as it should increase as soon as you add a new explanatory variable to the model. Numerous criteria have been proposed in the literature (see Table 6.1 for the formulas), which take into account the number of observations (the length of the series) and the number of explanatory variables. Among the most popular we have: Mallow's Cp, Akaike's criterion (AIC), Schwarz's criterion (BIC), the adjusted R-square etc.

Table 6.1: Formulas and Definitions for Model Fit Summary Statistics

Statistic	Definition or Formula
T	Number of observations
N	Number of parameters including the intercept
$\hat{\sigma}^2$	Estimate of error variance from fitting the full model
SST	Total sum of squares corrected for the mean for the dependent variable
SSE	Sum of square errors
ASE	$\frac{SSE}{T}$
MSE	$\frac{SSE}{T-N}$
R^2	$1 - \frac{SSE}{SST}$
$ADJRSQ$	$1 - \frac{(T-1)(1-R^2)}{T-N}$
AIC	$T\ln(\frac{SSE}{T}) + 2N$
$AICC$	$1 + \ln(\frac{SSE}{Y}) + \frac{2(N+1)}{T-N-2}$
C_p	$\frac{SSE}{\hat{\sigma}^2} + 2N - T$
BIC	$T\ln(\frac{SSE}{T}) + N\ln(T)$

6.3 Reduced Regression Methods

6.3.1 Subset Selection

With subset selection we retain only a subset of the variables, using some decision criterion, and eliminate the rest from the model. Least squares regression is then used to estimate the coefficients of the inputs that are retained. There are a number of different strategies for choosing the subset.

Best-Subset Selection

With N possible explanatory variables, the total number of candidate models is equal to 2^N and it is usually impossible to estimate all of them. But, if you focus on models with only $k < N$ variables, the number is 'smaller' and equal to $\frac{N!}{k!(N-k)!}$. Best subset regression finds for each $k \in (1, 2, 3, \ldots, N)$ the subset of size k that gives smallest residual sum of squares. An efficient algorithm, the *leaps and bounds* procedure proposed by Furnival and Wilson (1974), makes this feasible for N as large as 30 or 40.

Note that the best subset of size k, for example, may not include all the variables that was in the best subset of size $(k-1)$. The choice of k involves the tradeoff between bias and variance, and the more subjective desire for parsimony. There are a number of criteria that one may use; the most popular is to choose the smallest model that minimizes an estimate of the expected prediction error.

Forward, Backward and Stepwise Selection

Rather than search through all possible subsets (which becomes infeasible for N much larger than 40), we can seek a good path through them.

Forward selection (FS) begins with no variables in the model. For each of the independent variables, the algorithm calculates a given statistic S that reflects the variable's contribution to the model if it is included. The p-values for these S statistics are compared to a pre-specified threshold value. If no statistic has a significance level greater than the threshold value, the algorithm stops. Otherwise, it adds the variable that has the largest S statistic to the model. The algorithm then calculates S statistics again for the variables still remaining outside the model, and the evaluation process is repeated. Thus, variables are added one by one to the model until no remaining variable produces a significant S statistic. Once a variable is in the model, it stays. Like best-subset regression, Forward selection (FS) produces a sequence of models indexed by k, the subset size, which must be precised to the algorithm. Forward selection is a greedy algorithm, producing a nested sequence of models. In this sense it might seem sub-optimal compared to best-subset selection. However, there are several reasons why it might be preferred:

- Computational; for large N we cannot compute the best subset sequence, but we can always compute the forward stepwise sequence (even when $N >> T$).
- Statistical; a price is paid in variance for selecting the best subset of each size; forward stepwise is a more constrained search, and will have lower variance, but perhaps more bias.

Backward selection (BS) begins by calculating S statistics for a model which includes all of the independent variables. Then the variables are deleted from the model one by one until all the variables remaining in the model produce S statistics significant at a pre-specified level. At each step, the variable showing the smallest contribution to the model is deleted. Backward selection can only be used when $N < T$, while forward stepwise can always be used.

Forward stepwise selection (FSS) is a modification of Forward selection method and differs in that variables already in the model do not necessarily stay there. As in Forward-selection, variables are added one by one to the model, and the F statistic for a variable to be added must be significant at a pre-specified level (p-enter). After a variable is added, however, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce a S statistic significant at another pre-specified level

(p-stay). Only after this check is made and the necessary deletions are accomplished can another variable be added to the model. The stepwise process ends when none of the variables outside the model has an S statistic significant at the p-enter level and every variable in the model is significant at the p-stay level, or when the variable to be added to the model is the one just deleted from it.

With many candidate predictors, these algorithms require a lot of computation; even if accurate algorithms can exploit the QR decomposition for the current fit to rapidly establish the next candidate.

Forward-Stagewise Regression

Forward-stagewise regression (FSR) is even more constrained than forward-stepwise regression. It starts like forward-stepwise regression, with an intercept equal to 0, and centered predictors with coefficients initially all 0. At each step the algorithm identifies the variable most correlated with the current residual. It then computes the simple linear regression coefficient of the residual on this chosen variable, and then adds it to the current coefficient for that variable. This is continued till none of the variables have correlation with the residuals, i.e. the least-squares fit when $N > p$.

Unlike forward-stepwise regression, none of the other variables are adjusted when a term is added to the model. As a consequence, forward stagewise can take more than p steps to reach the least squares fit, and historically has been dismissed as being inefficient. It turns out that this 'slow fitting' can pay dividends in high-dimensional problems.

The GETS approach

The general-to-specific methodology has been advocated by David Hendry and his co-authors and discussed in some detail in a number of paper such as, e.g. Hendry (1995), Hendry (1997), Krolzig and Hendry (2001) and Campos et al. (2005), see also Brüggemann et al. (2003) for an application of this methodology to model reduction in VAR processes.

Starting from the experiments of Lovell (1983), Hoover and Perez (1999) proposed an algorithm for automatic econometric model selection and demonstrated that the general-to-specific procedure works well enough to be useful. Since this pioneering work, algorithms have considerably improved.

The salient features of the generic algorithm may be summarized as follows: 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 resulting specifications are acceptable using misspecification tests.

Doornik (2009) details the latest installment in the automated GETS methodology. The algorithm is based on the following main components[1]:

1. GUM: The general unrestricted model (GUM) is the starting point of the search. The GUM should be specified based on broad theoretical considerations to nest the local data generating process (LDGP).
2. Pre-Search: prior to specific selection, a pre-search lag reduction is implemented to remove insignificant lags, speeding up selection procedures and reducing the fraction of irrelevant variables selected (denoted the gauge of the selection process). Pre-search is only applied if the number of variables does not exceed the number of observations ($N < T$).
3. Search Paths: the algorithm uses a tree search to explore paths. Starting from the GUM, the algorithm removes the least significant variable as determined by the lowest absolute t-ratio. Each removal constitutes one branch of the tree. For every reduction, there is a unique subtree which is then followed; each removal is back-tested against the initial GUM using an F-test. If back-testing fails, no sub-nodes of this branch are considered (though different variants of this removal exist). Branches are followed

[1] The following description is directly taken from Hendry and Pretis (2011)

until no further variable can be removed at the pre-specified level of significance α. If no further variable can be removed, the model is considered to be terminal.

4. Diagnostic Testing: each terminal model is subjected to a range of diagnostic tests based on a separately chosen level of significance. These tests include tests for normality (based on skewness and kurtosis), heteroskedasticity (for constant variance using squares), the Chow test (for parameter constancy in different samples), and residual autocorrelation and autoregressive conditional heteroskedasticity. Parsimonious encompassing of the feasible general model by sub-models both ensures no significant loss of information during reductions, and maintains the null retention frequency of the algorithm close to α: see Diron (2008). Both congruence and encompassing are checked by the algorithm when each terminal model is reached after path searches, and it backtracks to find a valid less reduced earlier model on that path if any test fails. This repeated re-use of the original mis-specification tests as diagnostic checks on the validity of reductions does not affect their distributions (see Hendry and Krolzig (2003)).

5. Tiebreaker: as a result of the tree search, multiple valid terminal models can be found. The union of these terminal models is referred to as the terminal GUM. As a tiebreaker to select a unique model, the likelihood-based Schwarz information criterion (BIC) is used, though other methods are also applicable, and terminal models should be considered individually.

Castle et al. (2009) examine the efficacy of the general-to-specific modeling approach using a simulation framework. They compare twenty-one different model selection algorithms representing a diversity of approaches, including information criteria such as AIC and BIC, selection of a "portfolio" or best subset of models, general-to-specific algorithms, forward-stepwise regression approaches, Bayesian Model Averaging and inclusion of all variables.Their results show that GETS algorithms do as well or better than all others in over 90% of the experiments.

6.3.2 Shrinkage Methods

By retaining a subset of the predictors and discarding the rest, subset selection produces a model that is simpler (less variables) and easier to interpret. It might even have in some occasion lower prediction error than the full model. However, because it is a binary process, variables are either retained or discarded, it usually exhibits high variance, and so doesn't reduce the prediction error of the full model. Shrinkage methods are more continuous, and don't suffer as much from high variability.

Ridge Regression

Ridge regression, Hoerl and Kennard (1970a) and Hoerl and Kennard (1970b), shrinks the regression coefficients by imposing a penalty on their size. The ridge coefficients minimize a penalized residual sum of squares:

$$\beta^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{i=T} \left[y_i - \left(\beta_0 + \sum_{j=1}^{j=N} \beta_j X_{ij} \right) \right]^2 + \lambda \sum_{j=1}^{j=N} \beta_j^2 \right\}$$

Here $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage: the larger the value of λ, the greater the amount of shrinkage. The coefficients are shrunk toward zero (and each other). An equivalent way to write the ridge problem, which makes explicit the size constraint on the parameters, is:

$$\beta^{ridge} = \arg\min_{\beta} \left\{ \sum_{i=1}^{i=T} \left[y_i - \left(\beta_0 + \sum_{j=1}^{j=N} \beta_j X_{ij} \right) \right]^2 \right\}$$

Subject to

$$\sum_{j=1}^{j=N} \beta_j^2 \leq \alpha$$

There is a one-to-one correspondence between the parameters λ and α. When there are many correlated variables in a linear regression model, their coefficients can become poorly determined and exhibit high variance. By imposing a size constraint on the coefficients, this problem is alleviated: better estimation can be achieved on the average in terms of MSE with a little sacrifice of bias, and predictions can be improved overall. Note that the ridge solutions depend on the scaling of the inputs, and so one normally standardizes the inputs before solving the minimization problem.

LARS, LASSO and Elastic Net

Tibshirani (1996) introduced the Least Absolute Shrinkage and Selection Operator (LASSO), a penalized method similar to the ridge regression, but based on the L1-norm:

$$\beta^{LASSO} = \arg\min_{\beta} \left\{ \sum_{i=1}^{i=T} \left[y_i - \left(\beta_0 + \sum_{j=1}^{j=N} \beta_j X_{ij} \right) \right]^2 + \lambda \sum_{j=1}^{j=N} |\beta_j| \right\}$$

Or, equivalently:

$$\beta^{LASSO} = \arg\min_{\beta} \left\{ \sum_{i=1}^{i=T} \left[y_i - \left(\beta_0 + \sum_{j=1}^{j=N} \beta_j X_{ij} \right) \right]^2 \right\}$$

Subject to

$$\sum_{j=1}^{j=N} |\beta_j| \leq \alpha$$

Once more, the bound α is a tuning parameter: when α is large enough, the constraint has no effect and the solution is just the usual least squares regression. But, for smaller positive values of α, the solutions are shrunken versions of the least squares estimates and some of the β_j coefficients are equal to zero. Owing to the nature of the L_1-penalty, the LASSO does both continuous shrinkage and automatic variable selection simultaneously. Choosing α is like choosing the number of predictors to use in a regression model, and cross-validation is a good tool for estimating the best value for α. To solve for the LASSO estimator, Tibshirani (1996) used a combined quadratic programming method by observing that the LASSO constraint

$$\sum_{j=1}^{j=N} |\beta_j| \leq \alpha$$

is equivalent to combining 2^N linear constraints

$$\sum_{j=1}^{j=N} \omega_j \beta_j \leq \alpha,$$

with $\omega_j = \pm 1$. But the 'Least Angle Regression' (LARS) procedure, see Efron, Hastie, Johnstone and Tibshirani (2004), is a better approach that exploits the special structure of the LASSO problem, and provides an efficient way to compute the solutions simultaneously for all values of α. The least angle regression procedure follows the same general scheme than the Forward stepwise regression, but doesn't add a predictor fully into the model. The coefficient of that predictor is increased only until that predictor is no longer the one most correlated with the residual r. Then some other competing predictor is tested for inclusion in the model.

Algorithm 1. ***Least angle regression algorithm***

- *Start with all coefficients β_j equal to zero.*
- *Find the predictor X_j most correlated with Y.*
- *Increase the coefficient β_j in the direction of the sign of its correlation with Y. Take the residuals $r = y - \hat{y}$ and continue. Stop when some other predictor x_k has as much correlation with r as x_j has.*
- *Increase (β_j,β_k) in their joint least squares direction, until some other predictor x_l has as much correlation with the residual r.*
- *Continue until: all predictors are in the model.*

It can be shown that, with a slight modification, this procedure gives the entire path of LASSO solutions, as α varies from 0 to infinity. The modification needed is: if a non-zero coefficient hits zero, remove it from the active set of predictors and re-compute the joint direction. LARS and its variants are computationally efficient: an algorithm is available that requires only the same order of magnitude of computational effort as Ordinary Least Squares applied to the full set of covariates.

Although the LASSO has shown success in many situations, it has some limitations.

- In the $N > T$ case, the LASSO selects at most T variables before it saturates, because of the nature of the convex optimization problem. This seems to be a limiting feature for a variable selection method. Moreover, the LASSO is not well defined unless the bound α on the L_1-norm of the coefficients is smaller than a certain value.
- If there is a group of variables among which the pairwise correlations are very high, then the LASSO tends to select only one variable from the group and does not care which one is selected.
- For usual $T > N$ situations, if there are high correlations between predictors, it has been empirically observed that the prediction performance of the LASSO is dominated by ridge regression (Tibshirani, 1996).

The 2 first points make the LASSO an inappropriate variable selection method in some situations. Another regularization technique called the 'elastic net' has been proposed to solve these problems (Zou and Hastie (2005)). Similar to the LASSO, the elastic net simultaneously does automatic variable selection and continuous shrinkage, and it can select groups of correlated variables. It is like a stretchable fishing net that retains 'all the big fish' and is based on the following minimization problem, mixing L_1 and L_2 penalties.

$$\beta^{elastic} = \arg\min_{\beta} \left\{ \sum_{i=1}^{i=T} \left[y_i - \left(\beta_0 + \sum_{j=1}^{j=N} \beta_j X_{ij} \right) \right]^2 + \lambda \sum_{j=1}^{j=N} |\beta_j| + \gamma \sum_{j=1}^{j=N} \beta_j^2 \right\}$$

Simulation studies and real data examples show that the elastic net often outperforms the LASSO in terms of prediction accuracy.

A final regularisation method to discuss, was introduced by Zou (2006) and referred to as the adaptive LASSO (A-LASSO) estimator. For this estimator the L1-norms in the LASSO penalty are re-weighted. Zou shows that, if a reasonable initial estimator is available, under appropriate conditions, the A-LASSO correctly selects covariates with nonzero coefficients with probability converging to one, and that the estimators of nonzero coefficients have the same asymptotic distribution they would have if the zero coefficients were known in advance.

The optimisation problem now is:

$$\min_{\beta_N}\left\{\sum_{i=1}^{i=T}\left(y_i-\beta_0-\sum_{j=1}^{j=N}\beta_j X_{ij}\right)^2+\lambda\sum_{j=1}^{j=N}\widehat{w}_j\left|\beta_j\right|\right\}, \tag{6.1}$$

where $\widehat{w}_j = 1/|\widehat{\beta}_{init,i}|^{\gamma}$, $\widehat{\beta}_{init}$ is an initial estimator and $\gamma > 0$. Usually, the initial estimator is the LASSO estimator with the constraint parameter tuned in the usual way with a cross-validation scheme. Then, in the second stage cross-validation is again used to select the λ parameter in Equation (6.1).

Bridge regression

Frank and Friedman (1993) introduced the 'bridge regressions' based on the general penalization rule:

$$\beta^{ridge}=\arg\min_{\beta}\left\{\sum_{i=1}^{i=T}\left[y_i-\left(\beta_0+\sum_{j=1}^{j=N}\beta_j X_{ij}\right)\right]^2+\lambda\sum_{j=1}^{j=N}\left|\beta_j\right|^{\gamma}\right\}$$

Obviously, this procedure encompasses the ridge regression ($\gamma = 2$), the LASSO ($\gamma = 1$) and the subset selection ($\gamma = 0$). Fu (1998) proposed to select the value of the 2 parameters λ and γ via generalized cross-validation (GCV). Tibshirani (1996) and Fu (1998) compared the prediction performance of the LASSO, ridge and bridge regression (Frank and Friedman (1993)) and found that none of them uniformly dominates the other two.

6.3.3 Boosting

As an alternative to shrinkage regression, a number of researchers have developed methods that focus on the predictive power of individual regressors instead of considering all the N covariates together. This approach has led to a variety of alternative specification methods sometimes referred to collectively as "greedy methods". In this context, regressors are chosen sequentially based on their individual ability to explain the dependent variable. Perhaps the most widely known of such methods, developed in the machine learning literature, is "boosting" whose statistical properties have received considerable attention (Friedman (2001) and Friedman et al. (2000)). Boosting constructs a regression function by considering all regressors one by one in a simple regression setting, and successively selecting the best fitting ones. More details on boosting algorithms for linear models, and their theoretical properties can be found in Bühlmann (2006).

Bühlmann (2006) proves that boosting with the squared error loss, L_2 Boosting, is consistent for very high-dimensional linear models, where the number of predictor variables is allowed to grow essentially as fast as $O\left(e^T\right)$, assuming that the true underlying regression function is sparse in terms of the L1-norm of the regression coefficients. The use of an AIC-based method for tuning makes boosting computationally attractive since it is not required to run the algorithm multiple times for cross-validation. We closely follow the same algorithm as in Bühlmann (2006), which can be described as follows.

1. (Initialisation). Let $x_t = (x_{1t}, ..., x_{Nt})'$, $\mathbf{X} = (x_1, ..., x_N)$ and $\mathbf{e} = (e_1, ..., e_T)$. Define the least squares base procedure:

$$\widehat{g}_{\mathbf{X},\mathbf{e}}\left(\mathbf{x}_t\right)=\widehat{\delta}_{\widehat{s}}x_{\widehat{s}t},\quad \widehat{\delta}_i=\frac{\mathbf{e}'\mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i},\quad \widehat{s}=\min_{1\leq i\leq N}\left(\mathbf{e}-\widehat{\delta}_i\mathbf{x}_i\right)'\left(\mathbf{e}-\widehat{\delta}_i\mathbf{x}_i\right)$$

2. Given data $\mathbf{X}$ and $\mathbf{y} = (y_1, ..., y_t)'$, apply the base procedure to obtain $\widehat{g}^{(1)}_{\mathbf{X},\mathbf{y}}\left(\mathbf{x}_t\right)$. Set $\widehat{F}^{(1)}\left(\mathbf{x}_t\right) = \upsilon\widehat{g}^{(1)}_{\mathbf{X},\mathbf{y}}\left(\mathbf{x}_t\right)$, for some $\upsilon > 0$. Set $\widehat{s}^{(1)} = \widehat{s}$ and $m = 1$.

3. Compute residuals $\mathbf{e} = \mathbf{y} - \widehat{F}^{(m)}(\mathbf{X})$ where $\widehat{F}^{(m)}(\mathbf{X}) = (\widehat{F}^{(m)}(\mathbf{x}_1), ..., \widehat{F}^{(m)}(\mathbf{x}_T))'$ and fit the base procedure to the current residuals to obtain the fit $\widehat{g}^{(m+1)}_{\mathbf{X},\mathbf{e}}(\mathbf{x}_t)$ and $\widehat{s}^{(m)}$. Update

$$\widehat{F}^{(m+1)}(\mathbf{x}_t) = \widehat{F}^{(m)}(\mathbf{x}_t) + \upsilon \widehat{g}^{(m+1)}_{\mathbf{X},\mathbf{e}}(\mathbf{x}_t).$$

4. Increase the iteration index m by one and repeat step 3 until the stopping iteration M is achieved. The stopping iteration is given by

$$M = \min_{1 \leq m \leq m_{\max}} AIC_c(m),$$

for some predetermined large $m_{\max}$ where

$$\begin{aligned} AIC_c(m) &= \log(\sigma^2) + \frac{1 + tr(\mathcal{B}_m)/T}{1 - (tr(\mathcal{B}_m) + 2)/T} \\ \sigma^2 &= \frac{1}{T}(\mathbf{y} - \mathcal{B}_m\mathbf{y})'(\mathbf{y} - \mathcal{B}_m\mathbf{y}) \\ \mathcal{B}_m &= I - (I - \upsilon\mathcal{H}^{(\widehat{s}_m)})(I - \upsilon\mathcal{H}^{(\widehat{s}_{m-1})})...(I - \upsilon\mathcal{H}^{(\widehat{s}_1)}) \\ \mathcal{H}^{(j)} &= \frac{\mathbf{x}_j\mathbf{x}_j'}{\mathbf{x}_j'\mathbf{x}_j} \end{aligned}$$

$m_{\max} = 500$ and $\upsilon = \{0.1, 1\}$ values can be used as suggested in the literature.

6.4 Data reduction methods

Reduced regression algorithms can usually handle up to several dozens of indicators. But nowadays, the available economic databases contain thousands of potential explanatory variables. It is therefore necessary to make a first sort. In this work, even if expert knowledge remains the most important and relevant element in the selection of a sensible set of indicators, some well-known statistical techniques can be used to help the modeller. The resulting set of indicators can then be used in multiple regression techniques and can even be subject to reduction using for example one of the techniques presented in the previous section.

6.4.1 Principal Components and Factor Augmented Regressions

The most widely used class of data-rich forecasting methods is 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 forecasting equations which thereby becomes a standard forecasting equation as they only involve a few variables. The assumption is that the co-movements across the indicator variables x_t, where $x_i = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ is a vector of dimension $(N, 1)$, can be captured by a $(r, 1)$ vector of unobserved factors $F_t = (F_{1t}, F_{2t}, \ldots, F_{rt})$, i.e. $\tilde{x}_t = \Lambda' F_t + e_t$, 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, 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 the previous equation, 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 the equation, $\alpha' = \tilde{\alpha}'\Lambda\tilde{x}_t$, 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 the various factor methods relate 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 popularized by Stock and Watson (2002a) and Stock and Watson (2002b),

in the context of large temporal data sets, although the idea had been well established in the traditional multivariate statistical literature. The method of principal components (PC) is simple. Estimates of Λ and the factors F_t are obtained by solving:

$$V(r) = \min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{i=N} \sum_{t=1}^{t=T} \left(\tilde{x}_{it} - \lambda_t' F_t \right)^2,$$

where λ_t is a $(N, 1)$ vector of loadings that represent the N columns of $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$. One, non-unique, solution of this minimization program 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 computation of principal components, (see Bai (2003) and Stock and Watson (2002a)).

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) is designed to address this issue. 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, overall evidence suggests that static PC are a more effective, simple and robust technique for forecasting.

6.4.2 Partial Least Squares Regression

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 the sense that factors or components, which are linear combinations of the original regression variables, are used as regressors instead of the original variables. 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 Ξx_t, that give maximum covariance between y_t and Ξx_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.

Algorithm 2. ***Algorithm PLS***

1. *Set* $u_t = y_t$ *and* $v_{it} = x_{it}$, $i = 1, 2, \ldots, N$. *Set* $j = 1$.

[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.

2. *Determine the* $(N, 1)$ *vector of indicator variable weights or loadings* $w_j = (w_1, w_2, \ldots, w_{Nj})$ *by computing* $w_{ij} = \mathrm{Cov}(u_t, v_{it})$ *individual covariances:* $i = 1, 2, \ldots, N$*. Construct the* j*-th PLS factor by taking the linear combination given by* $w_j v_t$ *and denote this factor by* f_{jt}*.*
3. *Regress* u_t *and* v_{it}*,* $i = 1, 2, \ldots, N$ *on* f_{jt}*. Denote the residuals of these regressions by* $\tilde{u}_t$ *and* $\tilde{v}_{it}$ *respectively.*
4. *If* $j = k$ *stop, else set* $u_t = \tilde{u}_t$ *and* $v_{it}\tilde{v}_{it}$*,* $i = 1, 2, \ldots, N$ *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 forecast by regressing y_t on f_{jt}, $j = 1, 2, \ldots, k$. Helland (1988) and Helland (1990) provide a general description of the partial least squares (PLS) regression problem.

6.4.3 Cluster Analysis

Cluster Analysis and Time Series Analysis have both a very long history but surprisingly their paths scarcely crossed, at least until recently. In the last decade, when huge time series datasets became available, there has been an explosion of interest in mining time series data. Literally hundreds of papers have introduced new algorithms to index, classify, cluster and segment time series.

Cluster analysis (Everitt (1980)) aims to place objects into groups or clusters suggested by the data, not defined a priori, such that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar. Any clustering method is based on a measure to assess the 'similarity-dissimilarity' between 2 objects, a measure of the 'similarity-dissimilarity' between 2 clusters and on an aggregation strategy to build the clusters. Numerous strategies to build the clusters are usually available in standard statistical software: agglomerative hierarchical methods, optimal partitioning methods (K-means algorithm), divisive hierarchical methods, self organizing maps, etc. Hierarchical methods - which produce a set of imbricated clusterings - have the appealing property that they can be represented by a dendogram, in which the dissimilarity between two cases can be read from the height at which they join a same group.

Hundreds of distances are available for clustering survey data, of which the most popular is the Euclidean distance. But, as far as time series are concerned, the Euclidean distance, as well as other Minkowski type metrics, applied to 'raw' data may give very unintuitive and unexpected results. In particular, this distance is very sensitive to units and scales, does not allow comparing time series of different lengths, cannot handle local time shifting and is affected by the presence of noise and non-linearities (missing values, outliers, calendar effects etc.).

One way to solve these problems is to define new similarity measures: Dynamic Time Warping (Berndt and Clifford (1994)), Longest Common SubSequence (Das et al. (1997)) and Edit Distance on Real sequence (Chen et al. (2003)).

Another way is to preprocess the data, applying basic transformations: standardization, smoothing, interpolation, detrending etc. The distance functions discussed above are computed directly in the space of time series, eventually after preprocessing. Because computing these distance functions is generally computationally expensive, dimensionality reduction strategies are used. The basic idea is then to project the series into a space of distance-preserving transforms and to use only a small number of coefficients of the transform in the clustering. These high level representations of the time series often allow both to handle distortions in elegant ways and to achieve efficiency of the computations.

Several projection or decomposition techniques were proposed, some of them dealing explicitly with non-stationary time series:

- Autocorrelation Functions (ACF, PACF, IACF): Maballée et al. (1911), Wang and Wang (2000);
- Discrete Fourier Transforms (DFT): Agrawal et al. (1993), Maharaj (2002), Caiado et al. (2006);

- Discrete Wavelet Transforms using Daubechies or Haar basis, Cosine Wavelets: Huntala et al. (1999);
- Chebyshev Polynomials: Ng and Cai (2004);
- ARIMA coefficients : Piccolo (1990);
- Linear Predictive Coding of the Cepstrum (LPC): Kalpakis et al. (2001);
- Singular Value Decomposition through Principal Component Analysis: for example Korn et al. (1997), Cleveland (2004);
- Smooth Localized Complex Exponential model: Huang et al. (2004);
- Piecewise Linear Approximation: Morinaka et al. (2001);
- Piecewise Aggregate Approximation: Keogh et al. (2000);
- Adaptive Piecewise Constant Approximation: Keogh et al. (2001).

It has to be noted that some degree of supervised learning could be achieved here using cross-correlations or cross-spectra between the candidate variables and the response variable.

A third approach to quantifying similarity consists in projecting time series in the space of their probability distributions and therefore assumes that time series are stationary (Focardi (2001)). This projection allows to measure the distance between series of different lengths and with shapes that - though similar in their distributions - cannot be directly matched. It implements a more abstract notion of similarity than pattern matching. Natural distance functions in the space of probability distributions are the Kullback-Leibler divergence and the Chernoff information measures (Kakizawa et al. (1998)). Distances in the probability space can also be defined considering probabilistic models of the time series. Models based on Hidden Markov Models with an unknown number of states (Ge and Smyth (2000)) are quite popular.

Once the clusters have been determined, they are summarized selecting one or several representatives: 'medoids' (variables closed to the cluster center), principal components or common factors. Once more, at this stage some degree of supervised learning can be introduced using the target variable in the choice. Then, a 'traditional' variable selection method is used on this restricted set of potential explanatory variables.

Bibliography

Agrawal, R., Faloutsosa, C., Swami, A. (1993), Efficient Similarity Search in Sequence Databases, Lecture Notes in Computer Science, Springer Verlag, 730, 69-84.

Bai, J. (2003), Inferential Theory for Factor Models of Large Dimensions, Econometrica, 71, 135-173.

Berndt, D., Clifford, J. (1994), Using dynamic time warping to find patterns in time series, AAAI-94 Workshop on Knowledge Discovery in Databases.

Brüggemann R., Krolzig H.M. and Lütkepohl H. (2003), Comparison of Model Reduction Methods for VAR Processes, Technical Report 2003-W13, Nuffield College, University of Oxford.

Bühlmann, P. (2006), Boosting for High-Dimensional Linear Models, Annals of Statistics, 34(2), 599-583.

Caiado, J., Crato, N., Pena, D. (2006), A Periodogram-based Metric for Time Series Classification, Computational Statistics and Data Analysis, 50, 2668-2684.

Campos, J., Ericsson, N. R., Hendry, D. F. (2005), General-to-specific Modeling, An Overview and Selected Bibliography. In General-to-Specific Modelling, Campos, J., Ericsson, N. R., Hendry, D. F. editors. Edward Elgar Publishing.

Castle, J.F., Qin, X., Reed, W.R. (2009), How To Pick The Best Regression Equation: A Review And Comparison Of Model Selection Algorithms. Working Paper 13/2009. Department of Economics and Finance, College of Business and Economics, University of Canterbury, Christchurch, New Zealand.

Chen, L., Ozsu, M.T., Oria, V. (2003), Robust and efficient similarity search for moving object trajectories, Technical Report, CS-2003-30, School of Computer Science, University of Waterloo.

Cleveland, W.P. (2004), Stability and Consistency of Seasonally Adjusted Aggregates and Their Component Patterns, Studies in Nonlinear Dynamics and Econometrics, Vol. 8, 2, Article 15.

Das, G., Gunopulos, D., Mannila, H. (1997), Finding similar time series, Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, 88-100.

Doornik, J. A. (2008), Encompassing and automatic model selection. Oxford Bulletin of Economics and Statistics, 70, 915-925.

Doornik, J.A. (2009), Autometrics, in Castle, J.L. and Shephard, N. (2009), The Methodology and Practice of Econometrics, Oxford University Press, pp. 88-121.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004), Least Angle Regression. Annals of Statistics, Vol. 32 (2), pp. 407-499.

Everitt, B.S. (1980), Cluster Analysis, Second Edition, London: Heineman Educational Books Ltd.

Focardi, F. M. (2001), Clustering economic and financial time series: Exploring the existence of stable correlation conditions, Discussion Paper 2001-04, InterTek Group, Paris.

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000), The Generalized Dynamic Factor Model: Identification and Estimation, Review of Economics and Statistics, 82, 540-554.

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(471), 830-840.

Frank, E., Friedman, J. (1993), A Statistical View of Some Chemometrics Regression Tools, Technometrics, Vol. 35, No. 2 (May, 1993), pp. 109-135.

Friedman, J. (2001), Greedy Function Approximation: a Gradient Boosting Machine, Annals of Statistics, 29, 1189-1232.

Friedman, J., Hastie, T., Tibshirani, R. (2000), Additive Logistic Regression: a Statistical View of Boosting, Annals of Statistics, 28, 337-374.

Fu, W. (1998), Penalized regressions: the bridge vs the LASSO. Journal of Computational and Graphical Statistics, vol 7, no. 3, 397-416.

Furnival, G., Wilson, R. (1974), Regressions by Leaps and Bounds, Technometrics, Vol. 16, No. 4 (Nov.), pp. 499-511

Ge, X., Smyth, P. (2000), Deformable Markov Templates for Time Series Pattern Matching, Technical Report N.00-10, University of California at Irvine.

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, Scandinavian Journal of Statistics, 17, 97-114.

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.

Hendry, D. F., Krolzig, H.-M. (2003), New developments in automatic general-to-specific modelling. In Stigum, B. P. (ed.), Econometrics and the Philosophy of Economics, pp. 379-419. Princeton: Princeton University Press.

Hendry, D. F., Pretis, F. (2011), Anthropogenic influences on atmospheric CO2. Discussion Paper Series 584. University of Oxford, Department of Economics.

Hocking, R.R. (1976), The Analysis and Selection of Variables in Linear Regression, Biometrics, 32, 1-50.

Hoerl, A.E., Kennard, R.W. (1970a), Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12, 55-67.

Hoerl, A.E., Kennard, R.W. (1970b), Ridge Regression: Applications to Nonorthogonal Problems, Technometrics, 12, 69-82.

Hoover, K.D., Perez, S.J. (1999), Data Mining Reconsidered: Encompassing and the General-To-Specific Approach to Specification Search, Econometrics Journal, 2, 167-191.

Huang, H.Y., Ombao, H., Stoffer, D.S. (2004), Discrimination and Classification of Nonstationary Time Series Using the SLEX Model, Journal of the American Statistical Association, Vol. 99, 467, 763-774.

Huntala, Y., Karkkainen, J., Toivonen, H. (1999), Mining for similarities in aligned time series using wavelets. Data Mining and Knowledge Discovery: Theory, Tools, and Technology, SPIE Proceedings Series, Vol. 3695, pp 150-160.

Kalpakis, K., Gada, D., Puttagunta, V. (2001), Distance measures for effective clustering of ARIMA time-series, Proceedings of the IEEE Int'l Conference on Data Mining, San Jose, CA, 273-280.

Kakizawa, Y., Shumway, R.H., Taniguchi, M. (1998), Discrimination and Clustering for Multivariate Time Series, Journal of the American Statistical Association, Vol. 93, 441, 328-340.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. (2000), Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases, Knowledge and Information Systems, Vol. 3, 263-286.

Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. (2001), Locally adaptive dimensionality reduction for indexing large time series databases, Proceedings of ACM SIGMOD Conference on Management of Data, 151- 162.

Korn, F., Jagadish, H., Faloutsos, C. (1997), Efficiently supporting adhoc queries in large datasets of time sequences, Proceedings of the ACM SIGMOD Int'l Conference on Management of Data.

Krolzig, H.M., Hendry, D.F. (2001), Computer Automation of General-to-Specific Model Selection Procedures, Journal of Economic Dynamics and Control, 25(6-7), 831-866.

Lovell, M. C. (1983), Data mining. Review of Economics and Statistics, 65, 1-12.

Maballée, Colette et Berthe (1911), Classification of Time Series and Forecasting: The SiNCiD Method, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 3, 159-167.

Maharaj, E.A. (2002), Comparison of non-stationary time series in the frequency domain, Computational Statistics and Data Analysis, 40.

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.

Ng, S. (2013), Variable Selection in Predictive Regressions, Handbook of Economic Forecasting, Elsevier, Edited by Graham Elliott and Allan Timmermann, Volume 2, Part B, Pages 752-789.

Ng, R.T., Cai, Y. (2004), Indexing Spatio-Temporal Trajectories with Chebyshev Polynomials, Proceedings of SIGMOD 2004.

Piccolo, D. (1990), A distance measure for classifying ARIMA models, Journal of Time Series Analysis, 11.

Stock, J.H., Watson, M.W. (2002a), 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. (2002b), Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business & Economic Statistics, 20, 147-162.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the LASSO, Journal of Royal Statistical Society, Series B, 58, 267-288.

Wang, C., Wang, X. S. (2000), Supporting content-based searches on time series via approximation, Proceedings of the 12th Int'l Conference on Scientific and Statistical Database Management, Berlin, Germany, 69-81.

Wold H. (1982), Soft Modeling, The Basic Design and Some Extensions, in Systems Under Indirect Observation, Vol. 2, ed. by J oreskog, K.G. and Wold, H., North-Holland, Amsterdam.

Zou, H. (2006), The Adaptive LASSO and Its Oracle Properties, Journal of the American Statistical Association, 101(746), 1418-1429.

Zou, H., Hastie, T. (2005), Regularization and Variable Selection via the Elastic Net, Journal of the Royal Statistical Society, Series B: 301-320.

7 Model Selection, Model Specifications and a Typology of Rapid Estimates

Authors

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

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

Fabio Bachini
Istat
bacchini@istat.it

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

7.1 Introduction

In the absence of complete sample information, models -of one type or another- are used to produce 'rapid estimates'. In this chapter we review alternative modelling approaches, distinguish their different properties and consider their practical implementation. In so doing we seek to understand when and how different models can and should be applied.

We distinguish different classes of 'rapid estimate' according to the extent to which they rely on statistical and econometric techniques to fill the information gap left by incomplete sample information due to publication lags. Specifically, we delineate 'flash' estimates from 'nowcasts'. Nowcasts are based on a more extensive use of statistical and econometric methods. The distinction between flash estimates and nowcasts can become blurred. But flash estimates are presumed to focus on using regression methods to fill the information gap rather than directly estimate the target variable. In addition, flash estimates tend to use more aggregated data and econometric methods, and in essence can be viewed as using 'simpler' methods closer to national accounting principles.

There is always a trade-off between the timeliness and accuracy of rapid estimates. Estimates can always be produced more quickly by exploiting less hard information, but we might expect their quality to deteriorate as a result. The flow of data, and its arrival over time, are critical in defining a given flash estimate or nowcast. And it is important for users of flash estimates and nowcasts to be aware of the reliability of the estimates they are using; this requires information both on the models deployed and evaluation exercises on the quality of existing (historical) flash estimates and nowcasts. While GDP is the most relevant variable for which rapid estimates are produced, we should not forget that such estimates can also involve other relevant macro-economic variables, such as inflation, industrial production and employment.

Below, to help fix ideas, we do focus discussion on the production of flash estimates and nowcasts for quarterly GDP growth. However, this focus is, in general, without loss of generality. The production of early GDP estimates often involves consideration of higher-frequency data; and this does raise additional modelling issues due to the use of mixed-frequency data. We indicate special cases when appropriate.

The plan of this chapter is as follows. Section 7.2 reviews, in non-technical terms, model specifications used for the production of flash estimates. Section 7.3 considers nowcasting models. In both cases, focus is on describing the essence of the main workhorse models rather than providing an exhaustive review of specific models of which there are many. Section 7.4 then offers some generic practical advice about how one might approach the production of rapid estimates for Euro area GDP growth.

7.2 Flash estimation

Since the first quarter of 2003 Eurostat, under pressure to provide an estimate of Euro area quarterly GDP ahead of its first official estimate at about 65 days after the end of the reference quarter, has produced a rapid GDP estimate - its so-called Flash estimate - available at 45 days.

The UK and the US currently produce their first official estimates of GDP about 25 days after the end of the quarter. At present almost all European member states produce flash estimates of GDP within 45 days; see Table 7.1 below. Indeed Belgium and Spain currently produce estimates earlier at 30 days.

But this has not always been the case. Indeed there is always a pressure on statistical offices to speed up delivery of their estimates. Inevitably, with resource constraints impeding the production of earlier/higher frequency official quantitative surveys, this means relying increasingly on the production of flash estimates and nowcasts – that perhaps exploit available within quarter information on indicator variables but use statistical models to 'fill in the gaps' and produce more timely estimates of the variable of interest.

But there is an expected trade-off between the timeliness and accuracy of rapid estimates. Estimates can always be produced more quickly by exploiting less information; but, we might expect the quality of the estimates to deteriorate as a result. It is important to quantify this trade-off. The user can then decide what is an

acceptable degree of accuracy for a given time-scale. Traditionally, rapid estimates were produced using simple regression-based approaches that exploited available information on selected auxiliary 'indicators'. These are variables which bear a close relationship to the variable of interest, say national or Euro area GDP, but are made available more promptly than the data for which they stand as a proxy.

In practice, there are a large number of potential indicator variables, both quantitative and qualitative. These indicators are published at different leads and lags, relative to the quarter of interest. It is critical to understand and exploit this flow of indicator data within and after the end of the quarter of interest. Section 7.4 below provides more practical details. Econometricians call this forecasting with 'ragged edge' real-time data. That is, the aim is to predict current quarter values of the target variable with potentially missing (indicator) data for this quarter.

Different models involve different ways of linking the indicator variables to GDP. This can be done at a quarterly or monthly frequency. It is an empirical question which is most sensible. In this section we distinguish regression-based methods for the production of flash estimates from temporal disaggregation methods. Section 7.3 then turns to nowcasting models.

7.2.1 Regression-based production of flash estimates

Since current and lagged values of indicator variables, and lags of GDP itself, can plausibly help explain GDP one must consider carefully how one selects the indicator variables for the model used to explain GDP growth. The number of possible indicator variables can easily get very large. The problem is then to either 'select', in some sense, the 'best-fitting' indicators (these best-fitting indicators can be chosen on the basis of both a priori or objective in-sample performance criteria) or 'reduce' the set of indicator variables 'automatically'. Once this is done the models that can be used to nowcast GDP growth are estimable using classical statistical methods - there are no degrees of freedom constraints. Different models involve different ways of linking the indicator variables to GDP. This can be done at a quarterly, monthly or mixed frequency. It is an empirical question which is most sensible.

Here we review the regression based approach to producing flash estimates for quarterly GDP growth. Focus for now is on regressions with a small number of indicator variables. The approach adopted is designed to comply with the criterion that the models used to produce flash estimates should be credible to national accountants. We see this as ruling out processes with lengthy lags in exogenous variables, since it is difficult to defend a situation where an indicator is sharply influenced by some other variable up to six months or so ago.

The desire to produce clear models with short lags is reinforced by the fact that in many cases the data series we have available are although monthly, generally short in duration. This makes it more difficult to explore long run (or cointegration) properties of the data satisfactorily and models are often regression equations constructed with the dependent variable entering only in logarithmic differences.

The modelling framework requires only a one period ahead forecast. This means that there is no distinction between single equation and multivariate models. Regression-based flash estimates are produced as special cases of the following general regression equation expressed at the quarterly frequency ($t = 1, \cdots, T$ quarters):

$$\Delta y_t = c + \sum_{p}^{i=1} \alpha_i \Delta y_{t-i} + \sum_{p}^{i=0} \sum_{k}^{j=1} \beta_{ij} x_{t-i,j} + u_t; (t = 1, \ldots, T) \tag{7.1}$$

where Δy_t is the log of the dependent variable (we continue to focus for illustrative purposes on quarterly GDP growth), x_t is the j-th (quarterly) indicator variable ($j = 1, \cdots, k$) in logs when appropriate, c is an intercept, p the number of lags and u_t a disturbance. All indicator variables that enter (7.1), if necessary, are differenced until stationary. See Section 7.4 for more practical details on data transformation choices. We also note that for short horizons the forecasting performance from univariate nonlinear models is typically worse or not much

better; see Stock and Watson (1999) and Marcellino (2008). We therefore confine attention to simpler linear models.

It should be noted that contemporaneous quarterly values of the indicator variables are included in (7.1). This reflects the fact that these indicators by their nature are published ahead of the variables to which they are assumed to relate, even though they may relate to the same time period.

Selecting the indicator variable(s)

An important practical question, to which we return in Section 7.4 below, is how should one select the relevant indicator variables, $x_{t,j}$. Judgement and experience will often be crucial. For example if components of GDP are available ahead of the publication of the GDP data themselves it makes sense to use these as 'indicators'.

However, statistical criteria might be used also or instead. These can be attractive when there are a large number of potential indicator variables to choose from. One common approach is to consider various specifications of (7.1). Specifically, given k indicator variables and a given number of lags (p), for $t = 1, \ldots, T$, one might consider all possible combinations of (7.1) of the $(p + 1).k$ exogenous and p lagged endogenous variables thus generated. Since, however, this creates a very large possible number of regressions and bearing in mind the well-known benefits of parsimony in forecasting models, it can be wise to limit oneself to those equations containing no more than, say, two explanatory variables. There are ${}^{[(p+1)k+p]}C_2 + [(p+1)k+p]+1$ such equations. One can then 'automatically' select the preferred model using the Bayesian Information Criterion (BIC). We might then use this model, and its estimated coefficients from the sample $t = 1, \ldots, T$, and the quarter $T + 1$ values of the explanatory variables in the preferred model to nowcast Δy_{T+1}. Recall that the $T + 1$ values of the indicator variables are published ahead of the $T + 1$ values for $\Delta y_{T+1}$and can therefore be exploited when nowcasting.

Hendry and Hubrich (2011) consider in detail the value of disaggregated indicators when forecasting an aggregate. Consideration of disaggregate data quickly increases k. See also Lui and Mitchell (2013) for more details on how disaggregate indicator can be exploited when seeking to produce rapid estimates for an aggregated variable.

Mitchell (2009) and Mazzi et al. (2013) find qualitative business tendency survey data are particularly helpful when producing flash estimates in the run up to the recent recession. This points to temporal instabilities in the preferred indicator(s); i.e. some indicators may work well at some points in time, but perform poorly at others. This is another issue users should be aware of. The preferred indicator(s) may not remain the same over time.

Monthly Bridge Equations

Various methods are available to produce early estimates of a quarterly variable like GDP exploiting monthly indicator variables. One option, considered in Section 7.2.2 below, is to consider the problem as one of constructing monthly GDP estimates. The estimation of monthly GDP amounts to a temporal disaggregation or a distribution problem. A popular alternative is to use bridge equations, since this framework sits naturally with the current focus on (small k) regression-based flash estimates.

Bridging involves linking monthly data, typically released early in the quarter, with quarterly data like GDP; e.g., see Salazar and Weale (1999) and Baffigi et al. (2004). In effect a two-equation system is now used to nowcast Δy_{T+1}, with the second equation comprising the forecasting model for the monthly variable $x_{t,j}$. The errors between the two equations, at the underlying monthly frequency, are assumed orthogonal so that the equations are estimated separately. In common with much previous work, see Diron (2008), it is popular to consider simple AR models for $x_{t',j}$:

$$x_{t',j} = \sum_{i=1}^{p} \beta_i x_{t'-i,j} + e_{t',j} \tag{7.2}$$

where $t = 1, \cdots, T_m$ denotes the monthly data with $m = 3$ months in the quarter.

The regression model for Δy_t, (7.1), is therefore estimated (in-sample, $t = 1, ..., T$) using hard quarterly data on $x_{t,j}$. However, when wanting to estimate Δy_{T+1}since we may only have partial information on $x_{T+1,j}$ (for some indicator variables, j) the predicted values $\hat{x}_{T+1,j}$ from the AR model are used instead when producing flash estimates from (7.1). For example, when wanting to estimate GDP growth with only two months of data on industrial production available, the final month in the quarter is forecast using the AR model. This forecasted value is then combined with the two months of hard data to obtain $\hat{x}_{T+1,j}$.

Extensions to the regression based approach, as traditionally implemented, have been suggested. These include Castle and Hendry (2010) which can handle a large number of indicators, k. But these methods, because of their more sophisticated econometric content, are arguably better classified in our typology as relevant for the production of nowcasts rather than flash estimates.

Benchmark models

To evaluate the performance of any 'rapid estimate' – whether a flash estimate or a nowcast - it is important to have a benchmark. Ability to beat the benchmark, systematically over time, suggests that the model is of use. In economic forecasting the most popular benchmark, which proves surprisingly difficult to beat except perhaps within quarter, is the autoregressive forecast which involves setting $\beta_{ij} = 0$ in equation (7.1). Often it is assumed that $p = 1$. The random walk model sets $\alpha_1 = 1$. This model is robust to structural breaks (see Clements and Hendry (1998)). Both of these benchmark models condition on the previous quarter's GDP estimate but, unlike flash or nowcasting methods, do not exploit any within quarter information.

7.2.2 Temporal disaggregation based production of flash estimates

A conceptually distinct approach to produce flash estimates when using mixed-frequency data is to employ temporal disaggregation methods. These essentially involve considering the relationship, or regression, between x and y above at the higher (monthly) rather lower (quarterly) frequency.

An important issue for any user to decide is whether the temporal disaggregation is applied to data (e.g. GDP) in (perhaps logarithmic) levels or in logarithmic first differences or growth rates.

Consider the production of early estimates for (the level of) quarterly GDP. While, as discussed above, one can use within-quarter information contained in monthly, say, indicator variables and then deploy the regression-based approach, the regression based approach models the relationship at the quarterly frequency. It may well be that the relationship is better defined at the monthly frequency, such that unknown monthly GDP is explained by monthly movements in the indicator variables. However, since, historically at least, we observe quarterly values for GDP we can and should impose the temporal aggregation constraint that when estimated the monthly GDP values sum to the known quarterly GDP value.

Flash estimates can then be produced by taking rolling quarterly averages of the interpolated monthly values. The seminal approach to produce high frequency estimates (of the level of the variable of interest) with temporal constraints is Chow, G., and Lin, A. L. (1971). This approach is used widely in statistics and implementable in many software packages.

Chow, G., and Lin, A. L. (1971) worked out in a regression-based context the optimal (in a least squares sense) solution to the following univariate temporal disaggregation problem. Given a $T-$vector y_t of quarterly observations for a national accounts aggregate and the $3T \times k$ (as there are $m = 1, \cdots, 3$ months in the quarter, t) matrix X of monthly observations on k related series or indicators, the problem is to estimate the unknown monthly values contained in the $3T$-vector for monthly GDP $y_{t,m}$.

In its popular form the Chow-Lin approach imposes, at best, a very specific and restrictive ad hoc structure on the dynamics. *A priori* it assumes that they are adequately captured by giving a first-order autoregressive structure to the disturbance term in an otherwise static model relating the observed quarterly GDP data to

the indicators. This is unlikely to be flexible enough to model the data well, given both the persistence in macroeconomic data and their co-trending nature.

This has motivated the use of dynamic temporal disaggregation methods. As Mitchell et al. (2005) explain, these can be equivalently deployed both in regression form and in state-space form (as in Harvey and Pierse (1984)). (EUROMIND offers a recent application; see Frale et al. (2011).)

In either case, estimates of unobserved monthly GDP are computed using monthly information on the indicators, related to GDP either via a regression or a state-space model. Importantly the monthly GDP estimates, $y_{t,m}$, are constrained such that they satisfy the temporal disaggregation constraint relating them to observed quarterly GDP data y_t, such that:

$$y_t = \sum_{m=1}^{3} y_{t,m} \tag{7.3}$$

When the data are in the logarithms of the original time-series this temporal aggregation constraint is nonlinear: the sum of the logarithms is not the logarithm of the sum. One solution is to follow Proietti and Moauro (2006) and use an iterative algorithm to ensure the nonlinear aggregation constraint is met exactly.

Alternatively, interest might rest with the growth rates of monthly GDP and the temporal disaggregation constraint can then be applied in first difference form as in Mariano and Murasawa (2003).

7.3 Nowcasting

The production of nowcasts, as defined above, is explicitly more dependent on the estimation of (more sophisticated) econometric models. Again these nowcasting models commonly involve the use of mixed-frequency data.

We distinguish the following nowcasting methods.

7.3.1 Factor models

The factor-based approach can consider a large(r) set of indicator variables, k, than traditional regression based approaches and summarises their information in a small number of (unobserved) common factors. These factors are then used to help predict the variable of interest. While, as Eklund and Kapetanios (2008) review, there are many factor based forecasting approaches, the most popular is the static (principal components) approach – recently popularised by Stock and Watson (2002) but dating back to Rhodes (1937). This involves extracting up to $r < k$ principal components from the set of indicators, possibly stacked over time also, and then relating these to GDP growth via a linear regression. Statistical tests for the appropriate number, r, of factors can then be deployed.

The extraction of common factors which represent the 'underlying state of the economy' has a long tradition going back to Burns and Mitchell (1946). Factor based rapid estimates therefore might be related to 'composite indicators'; since the common factor is often considered as an example of composite indicator used to extrapolate the target variable.

Alternatives to principal components analysis or nonparametric spectral based alternatives (as used, for example, in the production of the 'Eurocoin' indicator; see Altissimo et al. (2010)) are identification and estimation of the factors using a parametric model. For example the state space approach is often used when the set of indicator variables is quite small (say < 12); e.g. see Stock and Watson (1989).

More recently dynamic factor models which explicitly accommodate mixed frequency and ragged-edge data have been used to nowcast GDP growth by Giannone et al. (2008), Brauning and Koopman (2014) and others.

7.3.2 MIDAS

Mixed-data sampling (MIDAS) regressions, developed by Ghysels et al. (2007), also provide a means of running regressions that allow the variable of interest and indicators to be sampled at different frequencies.

The basic MIDAS model for quarterly GDP growth, Δy_t, with a single indicator variable, x_t^m, assumed to be sampled three times a quarter so that $m = 3$, is

$$\Delta y_t = c + \beta B(L^{\frac{1}{m}})x_t^m + u_t^m$$

where the K-th order lag polynomial is given as $B(L^{\frac{1}{m}}) = \sum_{k=1}^{K} B_j L^{\frac{j}{m}}$ and $L^{\frac{j}{m}} x_t^m = x_{t-j,m}^m$.

The 'trick' of MIDAS regressions is to provide a parsimonious representation for this polynomial to avoid parameter proliferation, commonly using Exponential Almon lags.

But in macroeconomic applications differences in sampling frequency (e.g. monthly to quarterly) are actually often small, so the parameter proliferation which MIDAS seeks to avoid is often less acute.

Mixed frequency VAR models

Vector Auto Regressive (VAR) models are also increasingly used to produce nowcasts since the VAR model can accommodate mixed-frequency data. This is readily appreciated given that a VAR model, specified at the monthly frequency, can be caste in state-space form and the Kalman filter used to skip missing observations (as, for example, GDP is observed quarterly not monthly). A temporal aggregation constraint, as seen in (7.3) above, can then be imposed following Harvey and Pierse (1984) by augmenting the state-space model with a cumulator variable. Both classical (Mariano and Murasawa (2010)) and Bayesian (e.g. Schorfheide and Dongho (2014)) estimation algorithms might be deployed by the user.

7.3.3 Tests for model comparison

The discussion above, while itself deliberately parsimonious, already indicates that there are a large number of possible models that the user might use to produce their nowcasts. Model selection methods therefore provide one means to decide upon which model should be used. Indeed, Section 7.2.1. above has already indicated how model selection criteria like the BIC might be used. Automatic model selection algorithms, as seen in the Autometrics software package, also provide a means to arrive at a single model from many candidate models (see Castle and Hendry (2010)). Model comparisons are also commonly undertaken on an out-of-sample basis. This involves horse-racing different models over an out-of-sample window, ideally using real-time data vintages, so that the user can mimic real-time use of the competing models. The ranking of different model's nowcasts then provides helpful guidance to the user as to the preferred nowcasting model. In reality, it is often found that the preferred model changes over time; e.g. see Rossi (2013). That is, while one model may nowcast well over one sample period, another model nowcasts well over another sample period. See also Mitchell (2009) and Mazzi et al. (2013) for applications nowcasting UK and Euro area GDP growth.

7.3.4 Combination and rolling regressions

As a result it can be dangerous to rely on a single model, however carefully selected, when nowcasting. This is particularly so at times of structural change. This motivates the use of combination methods for the production of nowcasts. For a nowcasting application which compares combination and selection methods see Kuzin et al. (2013). They find that overall combination methods deliver more reliable and stable nowcasts, although results do vary by specification and sample etc. Combination offers a means of integrating out model uncertainty, in other words of insuring ourselves against having picked the wrong (regression) model. There is a considerable body of work that has found forecast combination to often work well; see Timmermann (2006) for a recent survey. Indeed equal weighting is often found to work as well as more complex (optimal - variance

weighted) alternatives (see Smith and Wallis (2009) and many confine attention to use of equal weights. To account for structural change, and parameter instability, an easy to apply approach is to follow Pesaran and Timmermann (2007), and others, and increase the set of models considered to consider nowcasting models estimated over not just the available sample period (which essentially amounts to use of an expanding window - since as more data arrive the sample period used for estimation increases), but a window of fixed length. While a simple means of picking up structural change, this method has been found to be effective.

7.3.5 Uncertainty and Instabilities

Mazzi et al. (2013) also argue that it is important to indicate the uncertainty of nowcasts by producing density nowcasts. Users then know how much confidence or weight to attach to a nowcast. They can also make probabilistic statements. They consider density nowcast combination as a means of improving the robustness of nowcasts to structural instabilities.

7.4 Practical empirical advice

In this empirical section, we survey - in generic terms - the practical problems one has to solve when building a flash estimate or a nowcast of, for example, Euro area GDP.

7.4.1 Why Euro area GDP - and which measure of Euro area GDP?

Why?

US GDP estimates are released about 30 days after the end of the reference quarter $(t + 30)$, while the first estimate of Euro area GDP is released around $t + 45$ days - like Japanese GDP. It is important for many users to have Euro area estimates to the same timescale as US estimates. Euro area GDP estimates are computed as the aggregation of national GDP data, and it would be difficult to accelerate the national production processes. Production of a flash estimate or a nowcast seems to be the only viable alternative.

Which measure of GDP?

In fact, seasonally and working-day adjusted GDP, which is published and commented on widely, is the most obvious choice. In order to facilitate comparisons across countries it is important to remove higher-frequency effects and 'noise' from the data. Therefore, we should like to have an estimate of the Euro area seasonally and working-day adjusted quarterly GDP growth rate, in chain-linked volumes, with the reference year currently set at 2000 (at 2000 exchange rates).

When?

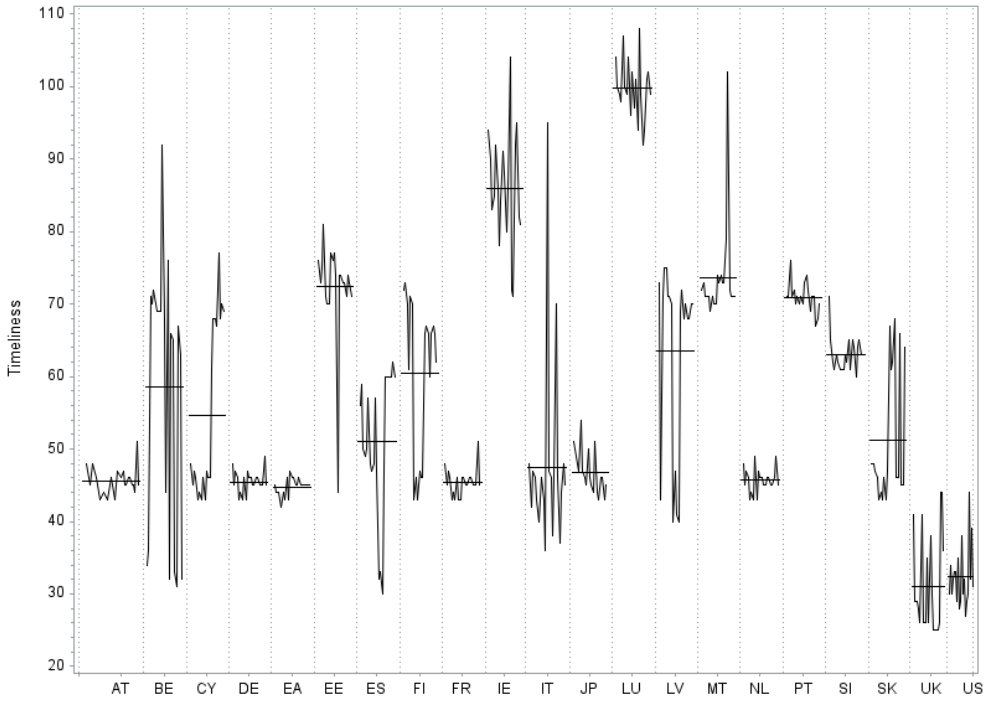
For no loss in accuracy, the earlier the estimate the better. But a widespread objective is to deliver the estimate by $t + 30$ days.

7.4.2 Timeliness

We define here the timeliness of an indicator as the number of days between the release date of the indicator (in fact its appearance in Eurostat's EuroInd database) and the end of the reference period it refers to. For example, if the GDP data for the first quarter of 2014 are released on May 13th, 2014 its timeliness is 43 days. Figure 7.1 and Table 7.1 compare the timeliness of GDP data for Euro area countries (Greece excluded), the Euro area as a whole, UK, US and Japan. From these data, it is clear that no Euro area GDP data are consistently available at $t+30$. This means that in the flash estimate or nowcasting model, e.g. equation (7.1) above, at best lagged values of GDP can be used.

In fact, 30 days after the end of a quarter, several important monthly indicators give us relevant information about the evolution of the economy for this quarter. For example, if we look at the production side, at least one or two months of the Industrial Production Index and the turnover indexes in industry, construction, wholesale and retail trade have already been published. These are therefore obvious indicators.

Figure 7.1: GDP Timeliness (in days) for various countries, on the period 2011Q1-2014Q1.



7.4.3 Flash estimate or nowcast?

As discussed above, a Flash estimate mimics as closely as possible the production process at the national statistical office. In our GDP example, a flash estimate for Euro area GDP should therefore be based on national GDP data alone, which are unfortunately not timely enough. We therefore have the following options:

- Forecast each country's GDP using an ARIMA model, and then aggregate these forecasts to obtain the Euro area estimate. Unfortunately, as these flash estimates will not be based on any timely (within quarter) data, it is likely that this simple strategy will miss turning points and other important dynamics.
- Forecast each country's GDP using a model like (7.1) which uses a small set of indicators, chosen using the user's judgement. Below in Section 7.4.4 we lay out popular choices for these indicators when producing flash estimates or nowcasts for GDP.
- Forecast each country's GDP using a model, of the sort reviewed in Sections 7.2 and 7.3, which uses a larger set of indicators chosen on the basis of statistical tests - or which extracts and then uses a common factor.
- Forecast only the bigger countries' GDP data using a model like (7.1); and estimate the remaining smaller country's GDP data using an ARIMA type model. This strategy may not involve much loss of accuracy as it is more important to have reliable estimates for Germany than Luxembourg, for example, when ultimately interested in Euro area GDP.

Table 7.1: GDP Timeliness (in days) for various countries.

Geo	11Q1	11Q2	11Q3	11Q4	12Q1	12Q2	12Q3	12Q4	13Q1	13Q2	13Q3	13Q4	14Q1	Mean
AT	43	47	46	47	45	45	46	46	45	45	44	51	45	45.8
BE	69	92	68	44	76	32	66	65	33	31	67	63	32	56.8
CY	43	47	46	46	61	68	68	67	77	68	70	69		60.8
DE	43	47	46	46	45	45	46	46	45	45	45	49	45	45.6
EE	76	77	74	44	74	74	73	73	71	74	73	71		71.2
ES	48	57	47	32	33	30	47	60	60	60	60	62	60	50.5
FI	43	47	46	46	66	67	66	60	66	67	66	62		58.5
FR	43	43	46	46	45	45	46	46	45	45	45	51	45	45.5
IE	91	88	80	86	104	72	71	92	95	82	81			85.6
IT	43	36	95	47	46	38	46	70	46	37	44	48	45	49.3
LU	96	102	97	101	94	108	99	92	94	101	102	99		98.8
LV	40	47	41	40	69	72	68	70	68	68	70	70		60.3
MT	70	74	73	74	73	73	79	102	72	71	71	71		75.3
NL	43	47	46	46	45	45	46	46	45	45	46	49	45	45.7
PT	71	70	73	74	72	69	71	71	67	68	70			70.5
SI	61	63	62	65	61	65	64	60	64	65	63	63		63.0
SK	43	47	67	61	62	68	46	46	66	45	45	64		55.0
EA	43	47	46	46	45	45	46	45	45	45	45	45	45	45.2
UK	35	26	38	30	25	25	25	25	25	26	44	44	36	31.0
JP	50	46	45	44	51	44	43	46	46	43	45			45.7
US	28	29	38	30	32	27	29	30	44	32	39	31	35	32.6

- Use the country level GDP data simultaneously to construct a nowcast for Euro area GDP. Chapter 15 of this handbook (Lui et al. (2017)) reviews 'aggregate', 'disaggregate' and multivariate approaches to nowcast an aggregate.

In our example, due to the lack of direct and timely information, it is likely that a nowcasting model would give better results than a flash estimate. That is, data availability at $t + 30$ days means one is forced to look beyond country-level GDP data when producing a rapid estimate for Euro area GDP. By looking at a variety of indicators, many of which are available monthly and within-quarter, one is moving away from the methodology used by the national accountants to estimate Euro area GDP. In this sense, one is producing a nowcast not a flash estimate.

7.4.4 Likely indicators for GDP

There are numerous possible monthly indicators for GDP that provide relevant within-quarter information. We delineate candidates below.

Industrial Production Index

For most Euro area countries, the industrial production index (IPI, B-D: Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply) for a given month is available before $t + 50$ days. The two exceptions are Austria and Belgium, which publish a bit later (see Table 7.2 and Figure 7.2). It means that for Euro area, and EA16 countries we already have at $t + 30$ days the estimates for two months of the quarter, and estimates for the first month of the quarter for the two remaining countries.

Figure 7.2: IPI Timeliness (in days) for various countries, on the period March 2013 to March 2014.

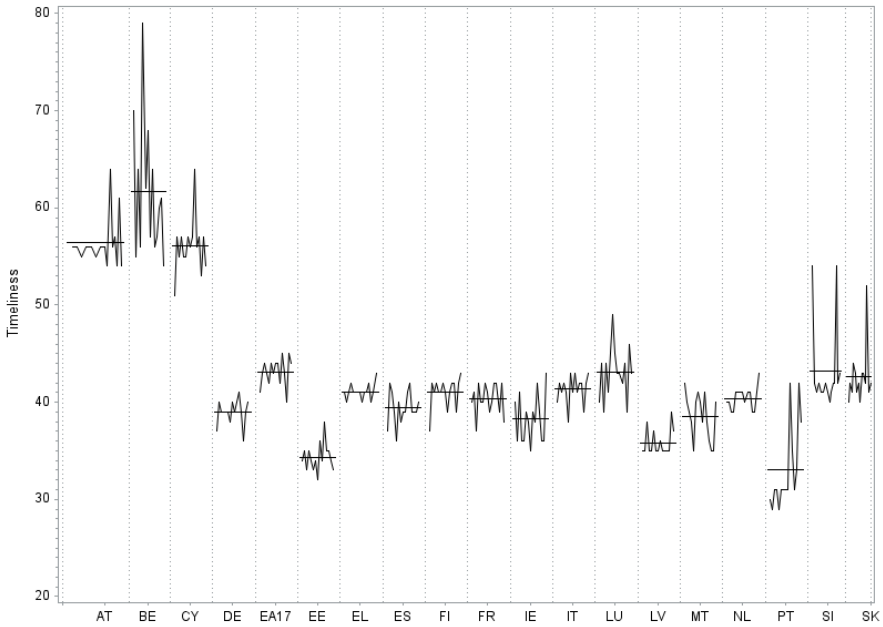
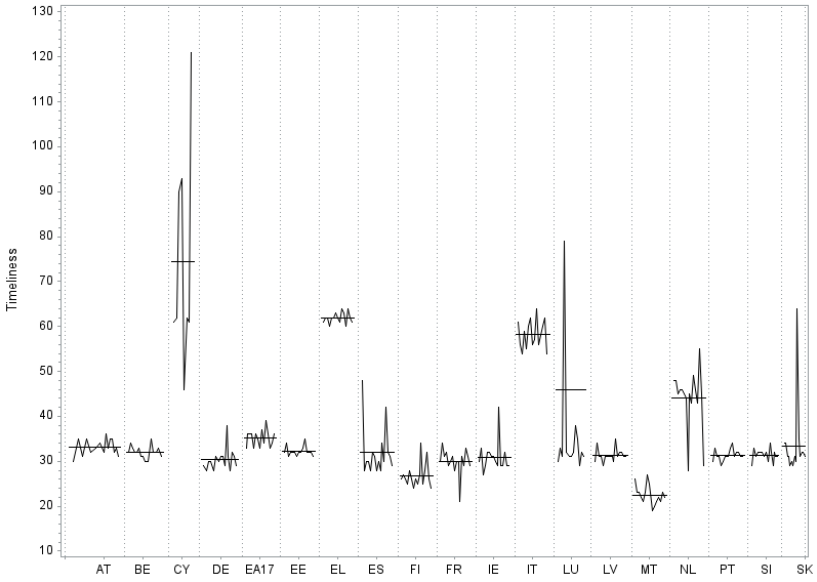


Table 7.2: IPI Timeliness (in days) for various countries.

Geo	13M3	13M4	13M5	13M6	13M7	13M8	13M9	13M10	13M11	13M12	14M1	14M2	14M3	Mean
AT	55	56	56	55	56	56	54	64	56	57	54	61	54	56.5
BE	55	64	56	79	62	68	57	64	56	57	60	61	54	61.0
CY	55	57	55	55	57	56	57	64	56	57	53	57	54	56.4
DE	39	39	39	39	38	40	39	40	41	39	36	39	40	39.1
EA17	44	43	42	44	43	44	44	42	45	43	40	45	44	43.3
EE	33	35	34	33	34	32	36	34	38	35	35	34	33	34.3
EL	41	42	41	41	41	41	40	41	41	42	40	41	43	41.2
ES	41	39	36	40	38	39	39	41	42	39	39	39	40	39.4
FI	41	42	41	41	42	41	39	41	42	42	39	42	43	41.2
FR	37	42	40	40	42	41	39	40	42	42	39	42	38	40.3
IE	41	36	36	39	38	35	39	38	42	39	36	36	43	38.3
IT	41	42	41	38	43	41	43	41	42	42	39	42	43	41.4
LU	39	44	41	45	49	45	43	43	42	44	39	46	43	43.3
LV	38	35	35	37	35	35	36	35	35	35	35	39	37	35.9
MT	39	38	35	40	41	40	38	41	38	36	35	35	40	38.2
NL	39	39	41	41	41	41	40	41	41	39	39	41	43	40.5
PT	31	31	29	31	31	31	31	42	35	31	33	42	38	33.5
SI	41	42	41	41	42	41	40	41	42	42	54	42	43	42.5
SK	41	44	43	41	42	40	43	43	42	52	41	42	43	42.8

Figure 7.3: Timeliness of the Retail Trade Turnover Index (in days) for various countries, on the period April 2013 to April 2014.



Retail Trade Turnover Index

The first estimates of the retail trade turnover index (DIT, G47) are published even earlier by most Euro area countries (see Table 7.3 and Figure 7.3) and for most of the Euro area countries at $t + 30$ days we have two months of data available within the quarter at $t + 30$ days.

Business Tendency Surveys

Business Tendency Surveys are very timely. The data for all Euro area countries (except Ireland) are delivered simultaneously by the DG ECFIN - at the end of the month under review or at the very beginning of the next month. All the data for the quarter are therefore available a few days, at the latest, after the end of the quarter of interest.

Business Tendency Surveys are 'pooled' surveys where businesses are asked about their appreciation of the current economic situation. These data, sometimes classified as 'soft data', should not be used in a flash estimation model but can be used for nowcasting. This is because their link to GDP is not direct, in contrast for example to the IPI which obviously forms part of GDP itself. Business tendency surveys, by contrast, are proxies for GDP and sub-component movements. These surveys cover all sectors of the economy; and a variety of questions are asked as seen from Table 7.5. Answers obtained from the surveys are typically aggregated in the form of 'balances'. Balances are constructed as the difference between the percentages of respondents giving positive and negative replies. EU and Euro area aggregates are computed on the basis of the national results. The balance series are then used to build composite indicators: First, for each surveyed sector, confidence indicators are computed as arithmetic means of answers (seasonally adjusted balances) to a selection of questions closely related to the reference variable they are supposed to track (e.g. industrial production for the industrial confidence indicator). Second, the results for the five surveyed sectors are aggregated into the Economic Sentiment Indicator, whose purpose is to track GDP growth at

Table 7.3: Timeliness of the Retail Trade Turnover Index (in days) for various countries.

Geo	13M4	13M5	13M6	13M7	13M8	13M9	13M10	13M11	13M12	14M1	14M2	14M3	14M4	Mean
AT	35	32	33	34	32	36	33	35	35	32	33	31		33.1
BE	32	32	33	31	31	30	30	35	32	32	33	31		32.1
CY	93		46		62	61			121					74.5
DE	30	28	31	30	31	31	29	38	31	28	32	31	29	30.4
EA17	36	33	36	35	33	37	34	39	36	33	34	36		35.1
EE	32	32	32	31	32	32	33	35	32	32	32	31		32.2
EL	60	62	62	63	62	61	64	63	60	64	62	61		61.9
ES	30	28	32	31	28	30	28	34	30	42	32	31	29	31.9
FI	25	28	26	24	26	25	27	34	25	28	32	26	24	26.8
FR	32	29	30	31	28	30	30	21	31	29	33	31	29	29.9
IE	29	32	32	31	31	30	29	42	29	29	32	29	29	30.9
IT	59	55	60	62	56	57	64	56	58	60	62	54		58.3
LU	79		32	31	31	32		38	35	29	32	31		35.7
LV	31	29	31	31	31	31	30	35	31	32	32	31	31	31.3
MT	22	21	23	27	25	21	19	20	21	22	21	23	22	22.4
NL	46	46	45	44	28	45	43	49	46	43	55	45	29	44.1
PT	31	29	30	31	31	31	33	34	31	32	32	31	31	31.3
SI	32	32	32	31	32	31	30	34	32	29	32	31	31	31.4
SK	31	29	30	29	31	30	64	35	31	32	32	31	31	33.4

Table 7.4: Timeliness of the Business Tendency Surveys (in days) for various countries and on the period June 2013 to May 2014.

Geo	13M6	13M7	13M8	13M9	13M10	13M11	13M12	14M1	14M2	14M3	14M4	14M5	Mean
All EA countries	-2	-1	-1	-3	-1	-2	9	-1	-1	1	-1	-3	-0.8

Member State, EU and Euro area level. Finally, the European Commission produces the factor model-based Business Climate Indicator, which uses the results of the industry survey and is designed to assess cyclical developments in the Euro area.

Financial indicators and other potential indicators

Once we move beyond consideration of indicators directly related to GDP, principally because they are a component of GDP or at least closely related to one, we open the door to a very large number of possible monthly indicators. Financial and nominal (price) indicators have a special status. On the one hand, they are often found to help explain movements in GDP, including in crisis periods and around turning-points. But, on the other hand, their predictive power is in practice often found to be temporary (see Rossi (2013)).

7.4.5 Do we have to transform the data?

The answer to this question is very clear: Yes, we usually have to transform the data, and for many good reasons. Usually, the raw indicator variables and the raw target variable do not have the same dynamics. As shown in Figure 7.4 and Figure 7.5,

- The similarity between European GDP and IPI is more evident when the data are handled as growth rates than levels.
- Running regressions, as in (7.1), between variables which have a unit root (i.e. a stochastic trend) raises the risk of spurious results. It is therefore important to difference variables in equations like (7.1) until they are stationary.
- The retail trade sector and the industrial sector do not have the same seasonal pattern nor the same trading-day effects. It is therefore often better to work on seasonally adjusted data.
- Outliers and breaks in the series can drastically affect the stability and fit of a model. But the series might not present the same outliers (a strike could affect a specific sector) and it can be helpful to correct the series from these outliers by removing them or by taking logarithms to minimise their effect.

Table 7.5: Variables covered in the Business Tendency Surveys

Type of survey	Monthly questions	Quarterly questions
Industry	Production, past 3 months	Factors limiting production
	Production, next 3 months	Production capacity, current
	Total order books	Months of production secured
	Export order books	Order books, past 3 months
	Stocks of finished products	Export order books, next 3 months
	Selling prices, next 3 months	Capacity utilisation
	Firm's employment, next 3 months	Competitive position, domestic market
		Competitive position, EU markets
		Competitive position, extra-EU markets
Construction	Building activity, past 3 months	Months of production secured
	Factors limiting building activity	
	Overall order books	
	Firm's employment, next 3 months	
	Selling prices, next 3 months	
Retail trade	Business activity, past 3 months	
	Business activity, next 3 months	
	Stocks of goods	
	Orders placed with suppliers, next 3 months	
	Firm's employment, next 3 months	
	Selling prices, next 3 months	
Services	Business situation, past 3 months	Factors limiting business
	Demand/Turnover, past 3 months	Potential increase in volume of activity
	Demand/Turnover, next 3 months	
	Firm's employment, past 3 months	
	Firm's employment, next 3 months	
	Selling prices, next 3 months	

Figure 7.4: Euro area GDP (in black) and IPI (in red) - seasonally and working-day adjusted figures.

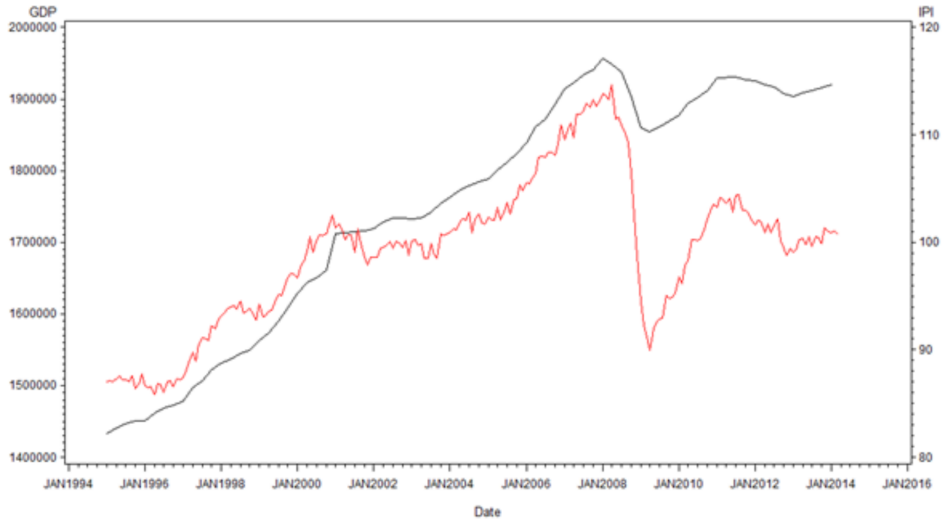
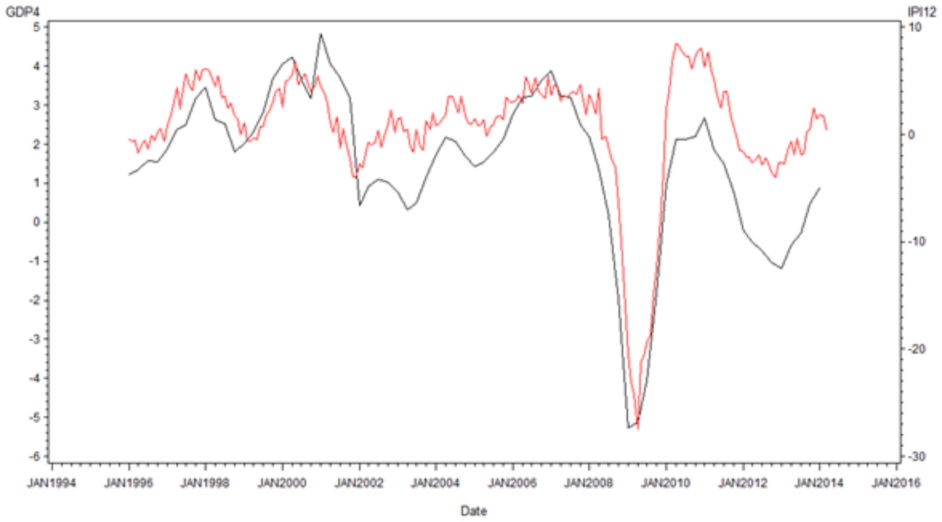


Figure 7.5: Euro area GDP (in black) and IPI (in red) - annual growth rates, seasonally and working-day adjusted figures.



7.4.6 When should we reduce the number of indicators?

We reviewed models in Sections 7.2 and 7.3 above that are operational when the number of indicators becomes too large for classical estimation of a simple regression model like equation (7.1). Of course, Bayesian shrinkage type estimators present an alternative.

But to give a simple example of how the number of indicators can quickly become quite large let us continue to think about producing rapid estimates for Euro area GDP growth. For a start, because we know GDP growth is a persistent time-series, we should look to use as regressors lagged GDP growth for the Euro area and for the 18 individual countries. Secondly, as considered in Section 7.4.4 above, we should use as indicators data on the IPI, the retail trade turnover index and the business survey sentiment indicator of these 19 geographical entities. We can also use the current and the first lag of these last three indicator variables. By this point we have $133(19 + 3 \times 2 \times 19)$ potential indicator variables and 2^{133} possible models.

This means that we may often wish to choose a smaller set of regressors – or combine estimates as discussed in Section 7.3.5. Chapter 6 of this handbook (Ladiray and Kapetanios (2017)) is dedicated to the selection methods you can use to guide your selection; but when using them, you must bare in mind some basic principles.

Building the model from economic reasoning leads to a parsimonious model and, if you are lucky, to good and stable estimates. This is usually a long 'trial and error' process that converges slowly. Hence the importance of trying out a methodology behind closed doors before publication. But economic reasoning is very helpful for the pre-selection of the potential explanatory variables. To take our previous example, we could restrict our set of potential variables to Germany, France, Italy, Spain, which together represent 80% of Euro area GDP.

We would then have 35 potential explanatory variables and 'only' 34 billion possible models. Starting from a large set of potential indicators, as discussed above, one can use a selection algorithm or a factor like method to compute 'new' variables that summarise the larger set of potential indicators. But, in any case, you should check the 'relevance' of the final model. For example, it can be hard to provide economic interpretation to some common 'factor' that is used to explain GDP growth.

7.4.7 What is a 'good' model for the production of rapid estimates?

This is always subjective, but we postulate the four following simple characteristics of a 'good' model:

- It should be as simple as possible.
- It should have a clear economic interpretation: it is important to understand the economic rationale behind the model to be able correctly to appreciate the quality of the results, and to interpret and communicate them.
- It should be robust and stable across time.
- It should provide 'accurate' estimates, as evaluated historically and in real-time on the basis of out-of-sample simulations – and as compared with a benchmark model. This uncertainty or error associated with the estimates should not be ignored – instead it should be acknowledged and reported.

When looking at these last two characteristics the user needs to deploy statistical criteria to evaluate the robustness and quality of the estimates. The only way to evaluate the quality of the model is to estimate it in 'real time' using a real-time database (the collection of Euro area GDP and explanatory indicators as they were published in the past). This will allow the user to compare the model estimates with the published Euro area GDP estimates, on the basis, for example, of the root mean square error or the mean absolute error statistics. If one does not have access to the data vintages, the best that can be done is to evaluate the performance of the model on the available data recursively adding a new observation to each out-of-sample simulation.

Three remarks conclude this section:

- It is always a good idea when looking for a model, to have a benchmark - a very simple model like an autoregressive model as in section 7.2.1 above - to evaluate the relative performance of the candidate models against.
- Keep in mind that one is trying to estimate a number that is itself revised. GDP data are revised many times, in the coming weeks and months. The user needs, therefore, to take a view on whether they are interested in a rapid estimate of the first estimate, the second estimate or indeed some later estimate. This can in turn have implications for what data (first release or second release etc.) are used to estimate the flash or nowcasting model (e.g. see Corradi et al. (2009)).
- Do not rely on a single model. Consult a variety of flash and nowcasting models of the type reviewed in Sections 7.2 and 7.3 above. And/or take a combination rapid estimate across a variety of models.

Bibliography

Altissimo F., Cristadoro R., Forni M., Lippi M. and Veronese G. (2010). New Eurocoin: Tracking Economic Growth in Real Time, Review of Economics and Statistics, 92(4), 1024-1034.

Baffigi A., Golinelli R. and Parigi G. (2004). Bridge models to forecast the Euro area GDP, International Journal of Forecasting, 20, 447-460.

Brauning F. and Koopman S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. International Journal of Forecasting 30 (3), 572âĂŞ584.

Measuring Business Cycles (New York, New York: National Bureau of Economic Research, 1946).

Castle J.L. and Hendry D.F. (2010). Nowcasting from disaggregates in the face of location shifts, Journal of Forecasting, 29(1-2), 200-214.

Chow G.C. and Lin A. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series, Review of Economics and Statistics, Vol. 53, pp. 372-375.

Corradi V., Fernandez A. and Swanson N.R. (2009). Information in the Revision Process of Real-Time Datasets, Journal of Business and Economic Statistics, 27(4), 455-467.

Clements M.P. and Hendry D.F. (1998). Forecasting Economic Time Series, Cambridge University Press: Cambridge.

Diron M. (2008). Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data, Journal of Forecasting, 27, 371-390.

Eklund J. and Kapetanios G. (2008). A review of forecasting techniques for large datasets, National Institute Economic Review, 203, 109-115.

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(2), 439-470.

Ghysels E., Sinko A. and Valkanov R. (2007). MIDAS Regressions: Further Results and New Directions, Econometric Reviews, 26 (1), 53-90.

Giannone D., Reichlin L. and Small D. (2008). Nowcasting: The real-time informational content of macroeconomic data, Journal of Monetary Economics, 55, 665-676.

Harvey A.C. and Pierse R.G (1984). Estimating Missing Observations in Economic Time Series, Journal of the American Statistical Association, Vol. 79, pp. 125-131.

Hendry D.F. and Hubrich K. (2011). Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate, Journal of Business and Economic Statistics, 29(2), 216-227.

Ladiray, D., Kapetanios, G. (2017). A review of parametric and non-parametric techniques for variable selection, in: Handbook on Rapid Estimates, ed. by D. Ladiray and G.L. Mazzi, Chapter 6, Eurostat.

Lui S., Mazzi G.L. and Mitchell J. (2017). Aggregate versus disaggregate approaches to construct rapid estimates, in: Handbook on Rapid Estimates, ed. by D. Ladiray and G.L. Mazzi, Chapter 15.

Lui S. and Mitchell J. (2013). Nowcasting quarterly euro area GDP growth using a global VAR model. In F. di Mauro & M. H. Pesaran (Eds.), The GVAR handbook: structure and applications of a macro model of the global economy for policy analysis. Oxford University Press.

Kuzin V., Marcellino M. and Schumacher C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries, Journal of Applied Econometrics, 28, 392-411.

Marcellino M. (2008). A linear benchmark for forecasting GDP growth and inflation?, Journal of Forecasting, 27, 305-340.

Mariano R.S. and Murasawa Y. (2003). A new coincident index of business cycles based on monthly and quarterly series, Journal of Applied Econometrics, 18(4), 427-443.

Mariano R.S. and Murasawa Y. (2010). A Coincident Index, Common Factors, and Monthly Real GDP, Oxford Bulletin of Economics and Statistics, 72(1), 27-46.

Mazzi G.L., Mitchell J. and Montana G. (2013). Density nowcasts and model combination: nowcasting Euro area GDP growth over the 2008-9 recession, Oxford Bulletin of Economics and Statistics, 76, 233-256.

Mitchell J. (2009). Where are we now? The UK recession and nowcasting GDP growth using statistical models, National Institute Economic Review, 209, 60-69.

Mitchell J., Smith R.J., Weale M.R., Wright S. and Salazar E.L. (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth, Economic Journal, 115, 108-129.

Pesaran M H. and Timmermann A. (2007). Selection of estimation window in the presence of breaks, Journal of Econometrics, 137, 134-161.

Proietti T. and Moauro F. (2006). Dynamic factor analysis with non-linear temporal aggregation constraints, Journal of the Royal Statistical Society: Series C, Vol. 55, pp. 281-300.

Rhodes E.C. (1937). The construction of an index of business activity, (with discussion), Journal of the Royal Statistical Society: Series C, 100, 18-66.

Rossi B. (2013). Advances in Forecasting under Model Instability, in: G. Elliott and A. Timmermann (eds.), Handbook of Economic Forecasting, Volume 2B, Elsevier Publications), 1203-1324.

Salazar E. and Weale M. (1999). Monthly data and short-term forecasting: an assessment of monthly data in a VAR model, Journal of Forecasting, 18, 447-462.

Schorfheide F. and Dongho S. (2014). Real-Time Forecasting with a Mixed-Frequency VAR, Journal of Business and Economic Statistics. (In Press)

Smith J. and Wallis K.F. (2009). A simple explanation of the forecast combination puzzle, Oxford Bulletin of Economics and Statistics. 71, 331-355.

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). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, Cointegration, causality and forecasting: a festschrift in honor of Clive W.J. Granger (editors Engle, R. and R. White). Oxford: OUP, pp. 1-44.

Stock J.H. and Watson M.W. (2002). Macroeconomic forecasting using diffusion indexes, Journal of Business and Economic Statistics, 20, 147-162.

Timmermann A. (2006). Forecast combinations. Handbook of Economic Forecasting, Volume 1 (editors Elliott, G., C W J. Granger and A. Timmermann). North-Holland: Amsterdam, pp. 135-196.

II Rapid Estimates based on Data Available at Different Frequencies

8 Multivariate Temporal Disaggregation

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Tommaso Proietti
University of Rome Tor Vergata
tommaso.proietti@uniroma2.it

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

8.1 Introduction

Temporal disaggregation methods play an important role for the estimation of short term economic indicators. This is certainly the case for a number of European countries (Eurostat (1999)), including France, Italy and Spain, whose national statistical institutes make extensive use of disaggregation methods, for constructing the quarterly national economic accounts using the annual figures and a set of indicators available at the quarterly frequency.

The methods that are currently in use are essentially univariate, a prominent role being played by the Chow-Lin procedure (Chow and Lin, 1971), which is based on a linear regression model on a set of indicators with stationary first order autoregressive errors. The main attraction of this approach is its simplicity. However, it suffers from a basic limitation: it does not use all the information available or it uses it in the wrong way, e.g. assuming the exogeneity of the indicators; furthermore, idiosyncratic movements in the indicators are transferred to the predictand via the regression function.

This chapter reviews the methods of multivariate temporal disaggregation proposed in the literature and evaluates their potential both from the theoretical and empirical standpoints. First of all we need to clarify what is intended by multivariate disaggregation methods. Essentially, an indirect estimation problem dealing with a system of time series variables, rather than a single one. The variables are typically connected by binding contemporaneous aggregation constraints. According to the Eurostat Handbook of Quarterly National Accounts (par. 6.35), *"multivariate models take into account the multivariate dimension of national accounts introducing contemporaneous accounting constraint in the estimation step in order to obtain estimates of national accounts aggregates which are coherent both temporally and in the accounting sense."*

Multivariate methods have the obvious advantage of using more information for the disaggregation task: the information has also not only an a priori nature, such as that originating from the contemporaneous definitional constraints between economic series, but also from entertaining a potentially large number of disaggregate time series concerning related phenomena.

This advantage has to be contrasted with the increased complexity and additional specification and estimation costs. As a result, the gains arising from multivariate disaggregation methods as opposed to the univariate approach are not clear cut, and a definite answer to this matter is primarily an empirical issue. A very fruitful approach, striking a good balance between model complexity and informativeness, is provided by dynamic factor models with mixed frequency data, which aim at extracting in a parsimonious way the information that is relevant for temporal disaggregation. The fundamental idea is to use all the potentially relevant information made available by the monthly time series, using simple and robust inferential methods at the same time. The fact that the series originate from the same economic environment should guarantee that relatively simple methods can estimate consistently the common features that are sufficient for the problem at hand.

The chapter has the following structure. Section 8.2 reviews the main multivariate approaches based on optimal or best linear prediction theory. The results are specialised to the case when both temporal and sectional constraints are enforced. We distinguish between two main classes of methods: those based on observation driven (reduced form) models and those based on parameter driven (structural time series) models. In section 8.3 we review the advances in multivariate temporal disaggregation using observation driven models.

Section 8.4 is dedicated to multivariate structural time series models and factor models. As hinted above, we believe that a good compromise between working in a data rich environment and model complexity is provided by an approach based on a dynamic factor model. The latter provides a computationally feasible way of summarising a wealth of monthly time series. Dynamic factor models have enjoyed a wide popularity recently, due to their capability of estimating a few latent factors that drive the co-movement of economic time series (see Forni et al. (2000); Forni et al. (2005); Stock and Watson (2002)).

Merging this approach with the state space methodology on temporal disaggregation provides a way of achieving the estimation of relevant quarterly information (the national accounts aggregates) at the monthly frequency. Among the results we obtain an estimate of monthly GDP and its components both from the supply

and the demand side consistent with the quarterly totals and that incorporated the cross-sectional aggregation constraints. This approach has been recently implemented by Proietti (2011a), Proietti (2011b) and Grassi et al. (2014).

As far as the statistical treatment is concerned, one of the distinctive traits is the adoption of the state space representation and the associated methods. The state space methodology offers the generality that is required to address a variety of inferential issues that have not been dealt with previously, such as the role of initial conditions and the definition of a suitable set of real time diagnostics on the quality of the disaggregation and revision histories that support model selection. Moreover, there are several computational advantages in the use of state space methods that are even more compelling in the multivariate framework (see Koopman and Durbin (2012)): the Kalman filter and smoother applied to the relevant state space form provides the Best Linear Unbiased Estimator (BLUE) of the disaggregate series. Also, essential diagnostic quantities, such as the innovations are not automatically available from the implementation of these methods within a classical regression framework.

We close the chapter by discussing how the practice of constructing chained-linked national accounts estimates modifies the way the sectional constraints are handled (section 8.5). Conclusions follow in section 8.6.

8.2 A review of the existing methods

There are several approaches available for the problem of multivariate temporal disaggregation, that have to be evaluated and assessed on the grounds of their realism, possibility of implementation in practice, identifiability of effects (see Palm and Nijman (1984)), and so forth. A possible taxonomy is the following:

1. Approaches based on reduced form time series models, such as VAR, VARMA, Error correction models.
2. Structural time series models (Harvey (1989); Moauro and Savio (2005)): these models are formulated in terms of unobserved components following given time series processes.
3. Semiparametric approaches, relating to the spline literature and local estimation under observation constraints.

Using a terminology due to Cox (1981) the first are also referred to as observation driven models, as the dynamic evolution is described in terms of “observables” such the past values and current and past innovations. The second are termed parameter driven as the evolution is a function of latent features, such as trends and cycles.

The third class can be viewed as a particular case of the former two and encompasses Denton’s multivariate benchmarking approach (Denton (1971)) and the polynomial method recently proposed by Hedhili and Trabelsi (2005). It usually requires the availability of a preliminary estimate of the disaggregate series. Benchmarking deals with combining information from different sources, characterised by different frequencies of observation and different reliability. Typically, there is a trade-off between these two features of the data, such that the less frequent (e.g. annual) series is more reliable and its values are considered as benchmarks. By benchmarking we aim at obtaining high-frequency estimates of a target variable that incorporates the more reliable information available from the benchmarks. See Grassi et al. (2014) and Dagum and Cholette (2006) for an overview. The literature on this topic is obviously overlapping with the topic of temporal disaggregation and has undergone several developments. Di Fonzo and Marini (2011) developed a simultaneous reconciliation procedure, arising as a multivariate extension of the modified Denton proportional first-difference benchmarking procedure (Denton (1971)). This solves the constrained optimization of an objective function according to which the proportionate difference between the benchmarked and the original series is as small as possible, subject to contemporaneous aggregation constraints. Di Fonzo and Marini, using sparse algorithms, show that a simultaneous adjustment of all the variables in the system is still feasible even when the system is very large.

8.2.1 A unified approach to multivariate temporal disaggregation

We provide a general analytic framework that encompasses the different multivariate temporal disaggregation methodologies currently in use. Let $\mathbf{y}$ denote a normally distributed random vector,

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_y),$$

whose mean is related to a set of explanatory variables, that we consider as exogenously determined. Typically, in our framework, the vector $\mathbf{y}$ is the stack of N unobserved time series of length n. The matrix $\mathbf{X}$ is known, though $\boldsymbol{\beta}$ may be unknown. We do not observe $\mathbf{y}$, but we observe

$$\mathbf{r} = \mathbf{R}\mathbf{y} + \boldsymbol{\epsilon},$$

with $\mathbf{R}$ denoting a suitable matrix of fixed coefficients and $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. The matrix $\mathbf{R}$ is typically a 'fat' matrix, i.e. its column dimension is much larger than the row dimension. In our set up, as we shall see in a moment, $\mathbf{R}$ is a matrix of temporal and contemporaneous aggregation weights, $\mathbf{r}$ is a vector of observed aggregate values. If the constraints are binding, then $\boldsymbol{\Sigma}_\epsilon$ is singular.

The conditional distribution of $\mathbf{y}$ given $\mathbf{r}$, assuming knowledge of $\boldsymbol{\beta}$ (and of $\boldsymbol{\Sigma}_y$, which is assumed throughout) is

$$\mathbf{y}|\mathbf{r}; \boldsymbol{\beta} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{B}(\mathbf{r} - \mathbf{R}\mathbf{X}\boldsymbol{\beta}), (\mathbf{I} - \mathbf{B}\mathbf{R})\boldsymbol{\Sigma}_y\right), \tag{8.1}$$

where

$$\mathbf{B} = \boldsymbol{\Sigma}_y\mathbf{R}'(\mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon)^{-1}$$

is the regression matrix of $\mathbf{y}$ upon $\mathbf{r}$.

Let us focus on the more realistic case when $\boldsymbol{\beta}$ is unknown, and let $\mathbf{b}$ denote a generic unbiased estimator that is linear in the observable $\mathbf{r}$, so that we can write, for a known $\mathbf{L}$ matrix

$$\mathbf{b} = \mathbf{L}'\mathbf{r}, \qquad \mathbf{L}'\mathbf{R}\mathbf{X} = \mathbf{I}.$$

The condition $\mathbf{L}'\mathbf{R}\mathbf{X} = \mathbf{I}$ implies $\mathrm{E}(\mathbf{b}) = \boldsymbol{\beta}$ (unbiasedness). The estimator mean square error matrix is $\mathrm{Var}(\mathbf{b}) = \mathbf{L}'(\mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon)\mathbf{L}$. The choice of $\mathbf{L}$ which minimises $\mathrm{Var}(\mathbf{b})$ is

$$\mathbf{L}' = (\mathbf{X}'\mathbf{R}'(\mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon)^{-1}\mathbf{R}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}'(\mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon)^{-1}.$$

This can be viewed as a generalisation of the Gauss-Markov theorem.

The estimator of the conditional mean of $\mathbf{y}$ given the aggregate data $\mathbf{r}$ is then

$$\widehat{\mathrm{E}}(\mathbf{y}|\mathbf{r}) = \mathbf{X}\mathbf{b} + \mathbf{B}(\mathbf{r} - \mathbf{R}\mathbf{X}\mathbf{b}), \tag{8.2}$$

and the conditional covariance matrix is

$$\widehat{\mathrm{Var}}(\mathbf{y}|\mathbf{r}) = \mathrm{E}[\mathrm{Var}(\mathbf{y}|\mathbf{r})] + \mathrm{Var}[\mathrm{E}(\mathbf{y}|\mathbf{r})]$$

where the variance and the expectation are computed with respect to the distribution of $\mathbf{b}$. The first term is $(\mathbf{I} - \mathbf{B}\mathbf{R})\boldsymbol{\Sigma}_y$, whereas $\mathrm{Var}[\mathrm{E}(\mathbf{y}|\mathbf{r})]$ is equal to the covariance matrix of $[(\mathbf{I} - \mathbf{B}\mathbf{R})\mathbf{X}(\mathbf{b} - \boldsymbol{\beta})]$ and thus

$$\widehat{\mathrm{Var}}(\mathbf{y}|\mathbf{r}) = (\mathbf{I} - \mathbf{B}\mathbf{R})\boldsymbol{\Sigma}_y + (\mathbf{I} - \mathbf{B}\mathbf{R})\mathbf{X}\mathbf{L}'(\mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon)\mathbf{L}\mathbf{X}'(\mathbf{I} - \mathbf{B}\mathbf{R})'$$

If the Gaussian assumption is removed, the previous result still provides the best linear unbiased predictor (BLUP) of $\mathbf{y}$. In particular (8.2) is known as the EBLUP (empirical BLUP) of $\mathbf{y}$ based on knowledge of $\mathbf{r}$ (see Robinson (1991)).

These expressions encompass the main multivariate disaggregation methods that belong to the BLUE ap-

proach (see Di Fonzo, 1990, 2002). In particular, $\mathbf{y}$ is a $nN \times 1$ vector stacking N time series of length n, generically denoted by $\mathbf{y}_i, i = \ldots, N$, i.e. it is

$$\text{vec}(\mathbf{Y}), \mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_N].$$

The matrix $\mathbf{Y}$ is not observable, due to temporal aggregation. Let $\mathbf{a}$ denote an $s \times 1$ aggregation vector, e.g. $\mathbf{a}' = [1, \ldots, 1]$ or $\mathbf{a}' = [0, \ldots, 0, 1]$ in the case of systematic sampling, and let $\mathbf{A}' = \mathbf{I}_m \otimes \mathbf{a}'$ denote the corresponding aggregation matrix which transforms $\mathbf{y}_i$ into the $M \times 1$ series $\mathbf{z}_i = \mathbf{A}'\mathbf{y}_i$. Thus, temporal aggregation implies that we observe the $m \times N$ matrix $\mathbf{Z} = \mathbf{A}'\mathbf{Y}$ in place of $\mathbf{Y}$.

Furthermore, the series are bound together by linear constraints. Most of the times these are deterministic, or binding, e.g. a national accounts identity, such as gross domestic product (GDP) by expenditure component adds up to the same total as GDP by branch of activity plus taxes less subsidies, or net exports equal total exports minus total imports. Contemporaneous aggregation constraints act on the columns of the matrix $\mathbf{Y}$ and thus can be expressed as $\mathbf{YC} = \mathbf{Q} + \mathbf{\Psi}$, where $\mathbf{C}$ is a known $N \times r$ aggregation matrix, $\mathbf{Q}$ is a known $n \times r$ matrix, and $\mathbf{\Psi}$ is a random matrix expressing the deviation from a deterministic constraint. We shall return later on the nature of the constraints that are relevant in our empirical investigation and their time varying nature, due to the chain linking of national account. The above formulation can entertain also the case of restrictions imposed by economic theory that pertain to the levels of the variables involved; for instance, a cointegrating relationship amounts to $\mathbf{c}_j'\mathbf{y}_t = q_j + \psi_{jt}$, where $\mathbf{c}_j$ is the j-th column of $\mathbf{C}$, $\mathbf{y}_t$ is the t-th row of $\mathbf{Y}$ and ψ_{jt} is a stationary random variable. Time invariant binding constraints determine a singularity in the distribution of the y's. This case is discussed in the next subsection.

Putting together the temporal and the cross-sectional constraints we obtain the observation equation

$$\mathbf{r} = \mathbf{R}\mathbf{y} + \boldsymbol{\epsilon},$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}_N \otimes \mathbf{A}' \\ \mathbf{C}' \otimes \mathbf{I}_m \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \text{vec}(\mathbf{Z}) \\ \text{vec}(\mathbf{Q}) \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \mathbf{0} \\ -\text{vec}(\mathbf{\Psi}) \end{bmatrix}.$$

In the sequel we assume knowledge of the $\mathbf{\Sigma}_y$ matrix; this is actually not realistic. The matrix is estimated from the observable $\mathbf{r}$ as well. Nevertheless, we do not account for the additional uncertainty arising because of its estimation. For this purpose we may apply the results in Doran, H.E. and Rambaldi, A.N. (1997), who derive a second-order approximation to the MSE and an estimator of the MSE correct to the same order.

8.2.2 Binding constraints and reparameterisations

When the cross-sectional constraints are time invariant and binding, $\mathbf{c}_j'\mathbf{y}_t = q_j$, $j = 1, \ldots, r, t = 1, \ldots, n$, the number of linearly independent time series is $N - r$, so that $\mathbf{y}_t$ has a singular normal distribution. The above formulae for the EBLUP and its MSE continue to hold, but with $\mathbf{\Sigma}_r^{-1}$ replaced by a Moore-Penrose pseudo-inverse, $\mathbf{\Sigma}_r^-$.

The system can be equivalently reparameterised in terms of a subset of $N - r$ linearly independent time series. To illustrate this let us start by writing $\mathbf{Q} = \mathbf{i}_n \otimes \mathbf{q}'$, where $\mathbf{q}$ is an $r \times 1$ vector and $\mathbf{i}_n$ is a vector of unit elements with length n. Thus, the constraint can be written

$$\mathbf{C}'\mathbf{y}_t = \mathbf{q}, t = 1, \ldots, n,$$

and partitioning $\mathbf{C}' = [\mathbf{C}_1', \mathbf{C}_2']$ so that $\mathbf{C}_2$ is $r \times r$ and nonsingular and $\mathbf{y}_t = [\mathbf{y}_{1t}', \mathbf{y}_{2t}']'$, where $\mathbf{y}_{2t}$ is a subset of r time series, we can express $\mathbf{y}_{2t}$ in terms of $\mathbf{y}_{1t}$:

$$\mathbf{y}_{2t} = -\mathbf{C}_2^{-1'}\mathbf{C}_1'\mathbf{y}_{1t} + \mathbf{C}_2^{-1'}\mathbf{q}.$$

In stack notation $\mathbf{y}_2 = -(\mathbf{C}_2^{-1'}\mathbf{C}_1' \otimes \mathbf{I}_n)\mathbf{y}_1 + (\mathbf{C}_2^{-1'}\mathbf{q} \otimes \mathbf{i}_n)$. This result implies that we can express the EBLUP

of $\mathbf{y}_2$ as a linear function of that of $\mathbf{y}_1$.

8.2.3 The BLUE approach

The best linear unbiased estimation approach assumes the availability of related series that are observed at the high frequency. The EBLUP can be viewed as a linear projection of the unobserved y's data on the available disaggregate information, corrected so as to take into account the temporal and sectional aggregation constraints. Each individual disaggregate time series is generated by the linear regression model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i,$$

so that using our notation, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \mathbf{u} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_y)$, $\mathbf{X}$ is a block diagonal matrix with diagonal blocks consisting of the $\mathbf{X}_i$ matrices, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_i', \ldots, \boldsymbol{\beta}_N']'$ and $\boldsymbol{\Sigma}_y$ is the covariance matrix of $\mathbf{u} = [\mathbf{u}_1', \ldots, \mathbf{u}_i', \ldots, \mathbf{u}_N']'$.

Assuming knowledge of $\boldsymbol{\Sigma}_y$, the vector $\boldsymbol{\beta}$ can be estimated by generalised least squares. Replacing the generating model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ into the observation equation, we obtain $\mathbf{r} = \mathbf{R}\mathbf{X}\boldsymbol{\beta} + \mathbf{R}\mathbf{u} + \boldsymbol{\epsilon}$. Denoting

$$\boldsymbol{\Sigma}_r = \mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}' + \boldsymbol{\Sigma}_\epsilon.$$

The GLS estimator of $\boldsymbol{\beta}$ is thus

$$\mathbf{b} = (\mathbf{X}'\mathbf{R}'\boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}'\boldsymbol{\Sigma}_r^{-1}\mathbf{r},$$

with $\mathrm{Var}(\mathbf{b}) = (\mathbf{X}'\mathbf{R}'\boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{X})^{-1}$. Hence, the EBLUP and its mean square error matrix are given respectively by:

$$\widehat{\mathrm{E}}(\mathbf{y}|\mathbf{r}) = \mathbf{X}\mathbf{b} + \mathbf{B}(\mathbf{r} - \mathbf{R}\mathbf{X}\mathbf{b}), \mathbf{B} = \boldsymbol{\Sigma}_y\mathbf{R}'\boldsymbol{\Sigma}_r^{-1},$$

$$\widehat{\mathrm{Var}}(\mathbf{y}|\mathbf{r}) = (\mathbf{I} - \mathbf{B}\mathbf{R})\boldsymbol{\Sigma}_y + (\mathbf{I} - \mathbf{B}\mathbf{R})\mathbf{X}(\mathbf{X}'\mathbf{R}'\boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{B}\mathbf{R})'.$$

Notice that for the GLS estimator the matrix $\mathbf{L}$ takes the expression: $\mathbf{L} = \boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{X}(\mathbf{X}'\mathbf{R}'\boldsymbol{\Sigma}_r^{-1}\mathbf{R}\mathbf{X})^{-1}$; obviously, $\mathbf{L}'\mathbf{R}\mathbf{X} = \mathbf{I}$. The expression for the EBLUP can be interpreted as a two stage procedure. The vector $\mathbf{X}\mathbf{b}$ is a preliminary estimate of the disaggregate series, which is corrected to comply with the observational constraints.

We finally observe that when all the constraints are binding, $\boldsymbol{\Sigma}_\epsilon = \mathbf{0}$ and $\boldsymbol{\Sigma}_r = \mathbf{R}\boldsymbol{\Sigma}_y\mathbf{R}'$, and the previous expressions simplify accordingly.

8.3 Observation driven models

We turn now to the crucial point concerning the specification of the underlying covariance matrix $\boldsymbol{\Sigma}_y$. The matrix plays an important role for the temporal and sectional smoothing of the disaggregation residuals $(\mathbf{r} - \mathbf{R}\mathbf{X}\mathbf{b})$.

We typically aim at achieving a specification that is identifiable from the aggregate data $\mathbf{r}$, which explains why the literature has focussed on two main assumptions:

White Noise disturbances $\boldsymbol{\Sigma}_y = \boldsymbol{\Omega} \otimes \mathbf{I}_n$, where $\boldsymbol{\Omega}$ is an $N \times N$ matrix. In particular, $\mathrm{E}(u_{it}, u_{js}) = \omega_{ij}I(t = s)$, where $I(\cdot)$ is the indicator function, taking the value 1 if the argument is true and 0 otherwise. When the constraints are fully binding, the matrix $\boldsymbol{\Omega}$ can be estimated using the OLS residuals of the regression of $\mathbf{A}\mathbf{y}_i$ on $\mathbf{A}\mathbf{X}_i$ and the computation of the EBLUP and its MSE matrix is considerably simplified.

Random Walk disturbances The disturbances are modelled as a multivariate random walk $u_{it} = u_{i,t-1} + \eta_{it}$, with contemporaneously correlated $\mathrm{E}(\eta_{it}, \eta_{js}) = \omega_{ij}I(t = s)$. If we let $\mathbf{M}$ denote the $n \times n$ matrix with generic element $M_{ts} = \min(t, s)$, then $\boldsymbol{\Sigma}_y = \boldsymbol{\Omega} \otimes \mathbf{M}$.

The random walk specification generalises the Fernández model (Fernández (1981)) to the multivariate case. When a preliminary estimate $\mathbf{y}^*$ is used in the place of $\mathbf{Xb}$, as it occurs in benchmarking, then the EBLUP is coincident with the estimate arising from Denton's additive first difference method. In the multivariate framework we have to account for the possibility that the series are cointegrated, due to the presence of long run equilibrium relationships between the constituent series.

8.3.1 Multivariate extension of the Chow-Lin method

The multivariate extension of the Chow-Lin method (Chow, G., and Lin, A. L. (1971)) is apparently straightforward as the univariate regression model with AR(1) errors can be translated into a multivariate regression with vector autoregressive errors. Letting $\mathbf{u}_t = [u_{1t}, \ldots, u_{Nt}]'$ denote the vector of disturbances at time t,

$$\mathbf{u}_t = \mathbf{\Phi}\mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathsf{N}(\mathbf{0}, \mathbf{\Omega}).$$

There are several issues that need to be explored as we move from the univariate to the multivariate framework, both from the theoretical and the empirical standpoints.

The first concerns the statistical identifiability of the matrix of autoregressive coefficient: as a result of temporal aggregation several parameters configurations can be observationally equivalent. This issue should be addressed using the methodology proposed in Casals et al. (2005) and Gómez and Aparicio-Pérez (2006).

One possibility is to assume $u_{it} = \rho_i u_{i,t-1} + \eta_{it}$, $\mathsf{E}(\eta_{it}, \eta_{js}) = \omega_{ij} I(t = s)$, i.e. an AR(1) process with series specific autoregressive coefficients. If further, it is assumed that $\rho_i = \rho$ for all i, then we face a system of seemingly unrelated time series, such that $\mathbf{\Sigma}_y = \mathbf{\Omega} \otimes \mathbf{M}$ and $M_{ts} = \rho^{|t-s|}/(1-\rho^2)$. The estimation is simplified considerably as the matrix $\mathbf{\Omega}$ can be concentrated out of the likelihood function, and the profile likelihood can be maximised with respect to $\rho \in (0, 1)$ by a grid search.

Another very interesting proposal is to use principal component analysis to derive a synthetic indicator from a large number of time series, to be used in the univariate Chow-Lin disaggregation; see Angelini et al. (2004).

8.3.2 The Multivariate Random Walk Model

In the national accounts framework the individual time series are linked by accounting identities. The problem was first considered by Rossi (1982) and Di Fonzo (1990). The cross-sectional constraints that are generated by the national accounts identities hold true when the series are expressed at current prices or at the prices of the previous year. For the chained volumes the constraints are no longer binding.

Proietti (2011b) proposed a general treatment of constraints that is valid both when they are binding and when they are stochastic and time varying. The former case is relevant for the temporal disaggregation of series expressed at current prices, whereas the second refers to the chained volumes. The presence of binding cross-sectional constraints implies that the variance-covariance matrix of the system of time series is singular. In the matrix treatment of the problem, considered in the previous section, and in Eurostat (1999), Annex A to ch. 11, the solution is provided in terms of the Moore-Penrose inverse of the system covariance matrix. An equivalent approach is to solve for the constraints and make them explicit. In particular, let $\mathbf{y}_{1t}$ denote an $N_1 \times 1$ set of unconstrained time series (eg. the value added of the six branches and Taxes less subsidies), then we can express N_2 series as linear combinations of $\mathbf{y}_{1t}$ as follows:

$$\mathbf{y}_{2t} = \mathbf{C}\mathbf{y}_{1t} + \boldsymbol{\epsilon}_t.$$

The MRW disaggregation method is based on the assumption that $\mathbf{y}_{1t}$ can be represented by a linear regression model with first order random walk errors:

$$\mathbf{y}_{1t} = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u}_t, \quad \mathbf{u}_t = \boldsymbol{\delta} + \mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathsf{NID}(\mathbf{0}, \mathbf{\Sigma}), \tag{8.3}$$

The model is completed by the measurement equation for the second block of time series, which is $\mathbf{y}_{2t} = \mathbf{C}\mathbf{y}_{1t} + \boldsymbol{\epsilon}_t$.

For instance, if $\mathbf{y}_{1t}$ is a vector containing seven time series, the value added of the six branches, and Taxes less subsidies, $\mathbf{y}_{2t}$ could be a vector with two elements, the first containing GDP at basic prices and the second containing GDP at market prices. In such case,

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The statistical treatment of this model, which entails using the state space methodology after the observational constraints due to temporal aggregation are taken into account, is given in Proietti (2011a) and Proietti (2011b).

8.3.3 The Multivariate Integrated Random Walk model

Due to the time series nature of the series involved, this specification is most suitable for the multivariate disaggregation of the national accounts series expressed at current prices. In this case the accounting identities are binding, and thus $\mathbf{y}_2 = \mathbf{C}\mathbf{y}_{1t}$. As a result, we only need to model the first component, $\mathbf{y}_{1t}$, and the estimate of the disaggregate time series for the second block is obtained deterministically from those of the first set of series. If an alternative direct disaggregate estimate of $\mathbf{y}_{2t}$ is available, then the estimates can be combined by some method, e.g. with weights proportional to the precision or the estimates, but this is a different issue.

The model that is considered in Proietti (2011b) is based on the state space representation of the multivariate Denton benchmarking technique (see, e.g., Di Fonzo and Marini (2005)); the model for $\mathbf{y}_t$ is obviously singular, due to the binding nature of the constraints, and rather than using a generalised inverse, we make the constraint explicit, by writing $\mathbf{y}_2 = \mathbf{C}\mathbf{y}_{1t}$ and disaggregating $\mathbf{y}_{1t}$ according to the model:

$$\begin{aligned} \mathbf{y}_{1t} &= \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u}_t, \\ \mathbf{u}_t &= \mathbf{u}_{t-1} + \boldsymbol{\delta}_{t-1}, \\ \boldsymbol{\delta}_t &= \boldsymbol{\delta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \mathsf{NID}(\mathbf{0}, \boldsymbol{\Sigma}), \end{aligned} \tag{8.4}$$

or, equivalently, $\Delta^2\mathbf{u}_t = \boldsymbol{\eta}_{t-1}$, i.e. an integrated random walk.

8.4 Multivariate structural (parameter driven) methods

The BLUE approach assumes the availability of a set of exogenous time series observed at the disaggregate frequency, which can be used to form a preliminary estimate. Harvey and Chung (2000) and Moauro and Savio (2005) have used multivariate structural time series models for the disaggregation problem, which overcomes some of the limitations of the regression based methods, namely the assumption of exogeneity of the indicators, see also Harvey (1989), sec 8.7.1, and the assumption that the indicators are measured without error. Durbin and Quenneville (1997) deal with the related topic of benchmarking by state space methods.

The specification of reduced rank components, such as common trends and cycles is also of interest. In particular, index models and factor models, such as Stock and Watson (1991) can be usefully adopted. Proietti and Moauro (2006) estimated monthly GDP for the U.S. and the euro area using a direct approach by formulating a dynamic factor model proposed by Stock and Watson (1989), specified in the logarithms of the original variables and at the monthly frequency. This poses a problem of temporal aggregation with a nonlinear observational constraint when quarterly time series are included (see also Proietti (2006)). The

main methodological contribution was to provide an exact solution to this problem, that hinges on conditional mode estimation by iteration of the extended Kalman filtering and smoothing equations.

The structural approach hinges upon the definition of unobserved components, e.g. trends, cycles and so forth, with a given dynamic structure. None of these references deals with the incorporation of binding cross-sectional constraints, e.g. national accounts identities.

Doran (1992) and Doran, H.E. and Rambaldi, A.N. (1997) consider the problem of incorporating time-varying constraints on the state vector of a state space model and handle it by by augmenting the measurement equation by artificial "observations". Doran (1992) estimates a Cobb-Douglas production function with time varying coefficient satisfying the constant returns to scale restriction whereas Doran, H.E. and Rambaldi, A.N. (1997) consider the estimation of constrained demand equations. Pandher (2002) and Pandher (2007) applies Doran's augmentation for modelling multivariate time series with multiple observational linear constraints. However, the constraint is not binding.

8.4.1 The state space methodology

Whatever is the approach taken, either parameter driven or observation driven, the state space methodology is very amenable to the statistical treatment of multivariate disaggregation methods. The computation of the EBLUP estimator of the disaggregate series is performed efficiently.

The exact initialisation of the multivariate disaggregation models is an issue of fundamental importance for the properties of the maximum likelihood estimates and for nesting exactly the regression methods within the more general class of dynamic regression models. This was illustrated in Proietti (2004), who however dealt only with the univariate case. The methodology that we propose is based in the implementation of the augmented Kalman filter and smoother approach set forth by de Jong (1991), see also de Jong, P., and Chu-Chun-Lin, S. (1994), suitably adapted to the specific problem at hand.

Also a new issue, not considered in the univariate case, is the initialization under partially nonstationary conditions, as in the presence of common trends. From the algorithmic standpoint one of the issues would be to adapt the Koopman and Durbin (2000) efficient multivariate Kalman filter and smoother to the augmented de Jong filter for handling the problem of temporal aggregation, dealt with in Proietti (2004). The Kalman filter and smoother also enable to define appropriate quality measures to be used for a systematic comparative assessment and evaluation process. The first comparison concerns the accuracy of the estimated disaggregate series, which can be evaluated essentially by two means:

- producing standard errors and constructing confidence intervals around the point estimates;
- documenting and assessing the process of revision of the estimates as the annual totals become available.

A set of efficient algorithms that are customized to perform the inference tasks taking into account the data generating process, that is temporal aggregation and ragged–edge data structures, are available. When missing data are present in a multivariate time series model, the best option is to use sequential processing (Anderson and Moore (1979); Koopman and Durbin (2000)). We also need to be able to entertain nonstationary as well as regression effects, which is done by assuming a diffuse prior on initial conditions and the effects of the explanatory variables (de Jong (1991)).

8.4.2 Temporal disaggregation using multivariate structural time series models: The SUTSE approach

The approach proposed by Moauro and Savio is based on the class of multivariate structural time series modes postulating the additive decomposition of the series into dynamic components representing trends, cycles, seasonality and the irregular component. The trend could be either a multivariate random walk or a

multivariate local linear trend process, integrated of order two. Non binding cross-sectional constraints can be handled by postulating that certain components, e.g. trends, are common.

Their SUTSE (seemingly unrelated time series equations) system is amenable since it does not assume any behavioural relationship between the series to be disaggregated and the set of related variables; the series to be disaggregated and the set of related series are simply affected by the same environment, so that they co-move and measure similar things, though none of them necessarily causes the other.

For non seasonal time series, the following model can the formulated at the higher frequency of observation (e.g. monthly):

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta} + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\eta). \end{aligned} \tag{8.5}$$

Here, $\boldsymbol{\mu}_t$ is the underlying level, evolving as a random walk with drift $\boldsymbol{\beta}$ (which could be allowed to vary over time), and $\boldsymbol{\epsilon}_t$ is the irregular component. Common trends can be handled by restricting $\boldsymbol{\mu}_t$ to be generated as follows: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\Theta}\boldsymbol{\mu}_t^\dagger$, where $\boldsymbol{\Theta}$ is an $N \times K$ matrix of loadings and $\boldsymbol{\mu}_t^\dagger$ is a K dimensional random walk.

The main problem with this approach is the curse of the dimensionality, that is the number of parameters is $O(N^2)$, and for very large dimensional systems, $N > 10$, estimation is prohibitive. The authors propose and EM approach to likelihood estimation which for small dimensional systems seems to work very well. Homogeneity restrictions (see Harvey (1989)) may also be plausible for certain applications.

8.4.3 Multivariate disaggregation by dynamic factor models

A multivariate disaggregation method should be feasible, transparent, capable of capturing the essence of the multidimensional information available and of enforcing the temporal and sectional aggregation constraints. A reasonable compromise is provided by the merge of the state space disaggregation methodology and dynamic factor models. Factor models enjoy widespread popularity, due to their capability of extracting the main economic signals from a massive number of time series.

Two main approaches have been considered that differ according to the dimension of the model (i.e. the number of time series considered) and the estimation method. A componentwise approach based on a small scale dynamic factor model, originally proposed by Stock and Watson (1989), has been proposed by Frale et al. (2011) for the construction of EUROMIND, a monthly indicator of the euro area gross domestic product. For each of the 11 components making up the decomposition of GDP by output and expenditure type a dynamic factor model is estimated with mixed frequency data and using the ragged-edge information structure arising from Eurostat publication schedule.

From the output side GDP is decomposed as follows[1]:

Label	*Value added of branch*	
A–B	Agriculture, hunting, forestry and fishing	+
C–D–E	Mining and quarrying, Manufacturing, Electricity, gas and water supply	+
F	Construction	+
G–H–I	Trade, transport and communication services	+
J–K	Financial services and business activities	+
L–P	Other services	=
	Total Gross Value Added	+
TIS	Taxes less subsidies on products	=
	GDP at market prices	

The breakdown of total GDP from the expenditure side is the following:

Label	*Component*	
FCE	Final consumption expenditure	+
GCF	Gross capital formation	+
EXP	Exports of goods and services	-
IMP	Imports of goods and services	=
	GDP at market prices	

The time series for the GDP components are quarterly and are available from the National Quarterly Accounts compiled by Eurostat. A set of coincident monthly indicators is also available for each component: for instance, in the case of the industry sector (C–D–E) we can consider the monthly index of industrial production and hours worked. For each individual GDP component, the monthly indicators and the quarterly GDP estimates are jointly modelled according to a single index dynamic factor model formulated at the monthly frequency and customised to handle ragged-edge data structures and temporal aggregation. The model parameters are estimated by maximum likelihood and signal extraction is performed by a suitable implementation of the Kalman filter and smoother handling sequential processing of the data as they become available according to their production schedule.

These methods yield estimates for the monthly GDP components in chain–linked volumes, with the year 2000 as the reference year. As it is well known, chain–linked volume measures are not consistent in aggregation, e.g. the sum of the value added of the six sectors plus taxes less subsidies would not deliver GDP at market prices, unless they are first expressed at the average prices of the previous years. Hence, to obtain our total monthly GDP estimates at market prices and chained volumes, Frale et al. (2011) resort to a multistep procedure that enforces the so–called annual overlap method advocated by the IMF (Bloem et al. (2001)).

Large scale factor models can be also considered. Grassi et al. (2014) estimate a generalization of EUROMIND, known as EUROMIND-C, which allows for the simultaneous calculation of monthly indicators of economic activity for the Euro area and its largest member states. EUROMIND-C is constructed based on more than 100 monthly time series and 55 quarterly national accounts series. The latter concern the decomposition of the gross domestic product according to the output and expenditure approaches, for the Euro area as a whole and for its four largest countries (Germany, France, Italy and Spain). The factor model postulates that the co-movements among the series are synthesized by a set of 16 common factors, representing the Euro area common trend, four country-specific factors and 11 component-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 that arises, also taking into consideration the ragged-edge structure of the dataset and the release schedule of the economic indicators. Computationally efficient algorithms, based on Jungbacker and Koopman (2012) and Jungbacker et al. (2011), have been implemented to avoid

[1] According to for example ISIC rev 3.1

the curse of dimensionality and to address specific issues related to the available data.

Unlike the original EUROMIND, the estimation of the monthly indicators of GDP and its components here is not carried out one component at a time, but simultaneously. In fact, the model includes all of the components of the breakdown of GDP, rather than performing a separate estimation for each different component (sector or expenditure type). Moreover, not only Euro area but country information is used, and country indicators of economic conditions are also produced. The model is specified at the monthly frequency and for the logarithms of the original series. The presence of national account aggregates, which are observed only quarterly, imposes constraints in relation to temporal aggregation. Furthermore, due to the logarithmic specification, a nonlinear smoothing algorithm has to be applied for the disaggregation of the quarterly national account aggregates. The corresponding state space representation is modified according to the observational constraints so as to handle the temporal aggregation and the ragged-edged data structure and missing values at the beginning of the sample period. Maximum likelihood estimation is carried out using the EM algorithm.

Proietti (2011a) estimated a large scale factor model for the Euro area is estimated with an ex post identification of the factors, the factors in this model are identified a priori and ascribed to a particular effect (country-specific factors, factors pertaining to a specific GDP component).

8.5 Chain-linking and contemporaneous aggregation constraints

The production of chained linked national accounts estimates has changed drastically the role of the contemporaneous aggregation constraints considered in section 8.4. Interestingly, due to the application of the annual overlap technique, exposed below, the constraints are not entirely lost, but they continue to hold after a transformation of the data that we call “dechaining”, aiming at expressing the chained values at the prices of the previous year.

The temporal disaggregation exercise is applied with respect to the quarterly chained volume measures of sectorial output and expenditure components produced by Eurostat. Currently these feature the year 1995 as the common reference year.

The Eurozone member states chain-link the quarterly data on an annual basis, i.e. the quarterly volume measures are expressed at the average prices of the previous years. The current situation is described in the Eurostat metadata available at http://ec.europa.eu/eurostat. Two alternative techniques are applied for annual chain-linking of quarterly data by the member countries: one quarter overlaps (Austria) and annual overlaps (other states). These are described in Bloem et al. (2001), chapter IX; the annual overlap technique, which implies compiling estimates for each quarter at the weighted annual average prices of the previous year, has the advantage of producing quarterly volume estimates that add up exactly to the corresponding annual aggregate. The annual overlap technique is also the method used by Eurostat in the imputation of the chain-linked volume measures of those countries for which no quarterly data at previous year’s prices are available.

As it is well known, chain-linking results in the loss of cross-sectional additivity (if the one quarter overlap is used also temporal additivity is lost and benchmarking techniques have to be employed in order to restore it). However, for the annual overlap, the disaggregated (monthly and quarterly) volume measures expressed at the prices of the previous year preserve both the temporal and cross-sectional additivity. If the estimates were expressed at current prices, then no consistency problem would arise, as the monthly disaggregated estimates would be perfectly additive.

The cross-sectional constraints can be enforced by a multistep procedure that de-chains the estimated monthly values, expressing them at the average prices of the previous year, and projects the estimates on the subspace of the constraints, as it will be described below. The dechaining procedure is in line with that advocated by the IMF manual (see Bloem et al. (2001)) and described in the sequel.

We start by indexing the month of the year by $j, j = 1, \ldots, 12$ and the year by m, $m = 1, \ldots, M = [n/12]$, so that the time index is written $t = j + 12m, t = 1, \ldots, n$.

For a particular estimated monthly time series let us denote by Y_{jm} the value at current prices of month j in year m, $Y_{.m} = \sum_j Y_{jm}$ the annual total, $\bar{Y}_m = Y_{.m}/12$ the annual average (the annual values are available from the national accounts). The chain-linked volume estimate with reference year b (the year 1995 in our case) will be denoted $\hat{Y}_{jm}^{(b)}$. The temporal disaggregation methods described in the previous section are applied to the quarterly chained-linked volume series with reference year b and yield estimates that add up to the quarterly and annual totals (temporal consistency), but are not additive in a horizontal (that is cross-sectional) sense.

The following multistep procedure enables the computation of volume measures expressed at the prices of the previous year that are additive.

1. *Dechaining* :

   a) Transform the monthly estimates into Laspeyres type quantity indices with reference year b (volumes are evaluated at year b average prices), by computing

   $$I_{jm}^{(b)} = \frac{\hat{Y}_{jm}^{(b)}}{Y_{.b}}, j = 1, \ldots, 12, m = 1, \ldots, M,$$

   where the denominator is the annual total of year b at current prices. In our case $b = 1$ (year 1 is the calendar year 1995).

   b) Change the reference year to $m = 2$, the second year of the series (1996 in our case), by computing:

   $$I_{jm}^{(2)} = \frac{I_{jm}^{(b)}}{\bar{I}_2^{(b)}}, j = 1, \ldots, 12, m = 1, \ldots, M,$$

   where $\bar{I}_2^{(b)} = \sum_j I_{j2}^{(b)}/12$ is the average quantity index for year 2.

   c) Transform the quantity indices for year $m = 2, 3 \ldots, M$ into indices with reference year $m-1$ (the previous year), by rescaling $I_{jm}^{(2)}$ as follows:

   $$I_{jm}^{(m-1)} = \frac{I_{jm}^{(2)}}{\bar{I}_{m-1}^{(2)}}, j = 1, \ldots, 12, m = 2, \ldots, M,$$

   where

   $$\bar{I}_{m-1}^{(2)} = \begin{cases} \frac{1}{12}\sum_j I_{j,m-1}^{(2)}, & m = 3, \ldots, M \\ \frac{Y_{.2}^{(b)}}{Y_{.b}}, & m = 2 \end{cases}$$

   d) Compute the series at the average prices of the previous year as:

   $$\hat{Y}_{jm}^{(m-1)} = I_{jm}^{(m-1)} \bar{Y}_{m-1}, \quad j = 1, \ldots, 12, m = 2, \ldots, M,$$

2. *Aggregation step*: Let $\mathbf{Y}_t$ denote the disaggregate time series expressed at the average prices of the previous year (we suppress reference to the previous year for simplicity). Using the original estimates and the dechained procedure we can assume that, at least approximately,

   $$\mathbf{Y}_t \sim \mathsf{N}(\hat{\mathbf{Y}}_t, \hat{\mathbf{V}}_t), t = 1, \ldots, n,$$

   where the first and second moments are given by the sequential constrained estimates produced by the Kalman filter and smoother outlined in the previous section, modified to take into account the dechaining

procedure[2]. If the r cross-sectional constraints are expressed as

$$\mathbf{R}\mathbf{Y}_t = \mathbf{r}$$

where $\mathbf{R}$ is an $r \times N_1$ matrix, the modified estimate that comply with those constraints and their MSE matrix are given respectively by

$$\begin{aligned} \tilde{\mathbf{Y}}_t &= \hat{\mathbf{Y}}_t + \hat{\mathbf{V}}_t\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_t\mathbf{R}')^{-1}(\mathbf{r} - \mathbf{R}\hat{\mathbf{Y}}_t) \\ \tilde{\mathbf{V}}_t &= \hat{\mathbf{V}}_t - \hat{\mathbf{V}}_t\mathbf{R}'(\mathbf{R}\hat{\mathbf{V}}_t\mathbf{R}')^{-1}\mathbf{R}\hat{\mathbf{V}}_t \end{aligned}$$

The new balanced estimates are now ready to be expressed at the average prices of year b.

3. *Chain-linking* (annual overlap):

   a) Convert the aggregated volume measures into Laspeyres-type quantity indices with respect to the previous year:

   $$\mathcal{I}_{jm}^{(m-1)} = \frac{\tilde{Y}_{jm}^{(m-1)}}{\bar{Y}_{m-1}}, \quad j = 1, \ldots, 12, m = 2, \ldots, M,$$

   where $\bar{Y}_{m-1} = \sum_j Y_{j,m-1}/12$ is the average GDP of the previous year at current prices.

   b) Chain-link the indices using the recursive formula (the first year is the reference year):

   $$\mathcal{I}_{jm}^{(1)} = \mathcal{I}_{jm}^{(m-1)}\bar{\mathcal{I}}_{m-1}^{(b)}, \quad j = 1, \ldots, 12, m = 2, \ldots, M,$$

   where

   $$\bar{\mathcal{I}}_{m-1}^{(1)} = \frac{1}{12}\sum_j \mathcal{I}_{j,m-1}^{(b)}.$$

   c) If $b > 1$ then change the reference year to year b:

   $$\mathcal{I}_{jm}^{(b)} = \frac{\mathcal{I}_{jm}^{(1)}}{\bar{\mathcal{I}}_b^{(1)}} \quad j = 1, \ldots, 12, m = 2, \ldots, M.$$

   (This is not needed in our case, since $b = 1$).

   d) Compute the chain-linked volume series with reference year b:

   $$\tilde{Y}_{jm}^{(b)} = \mathcal{I}_{jm}^{(b)}\bar{Y}_b \quad j = 1, \ldots, 12, m = 2, \ldots, M,$$

   where $\bar{Y}_b = \frac{1}{12}\sum_j Y_{jb}$ is the value of GDP (at basic or market prices) at current prices of the reference year.

The multistep procedure just described enables to obtain monthly estimates in volume such that the values $\tilde{Y}_{jm}^{(m-1)}$ expressed at the average prices of the previous year add up to their quarterly and annual totals published by Eurostat and are consistent with the contemporaneous aggregation constraints. Moreover, the chain-linked volumes $\tilde{Y}_{jm}^{(b)}$ with reference year b are temporally consistent, but are not horizontally consistent (cross-sectional additivity cannot be retained).

[2] In particular, if $\mathbf{D}_t$ is a diagonal matrix containing the dechaining coefficients that allow to express the chained estimates at the average prices of the previous year, and $\hat{\mathbf{Y}}_t^d$ are the chained estimates, with estimation error mean square matrix $\hat{\mathbf{V}}_t^d$, computed by the Kalman filter and smoother, then $\hat{\mathbf{Y}}_t = \mathbf{D}_t\hat{\mathbf{Y}}_t^d$ and $\hat{\mathbf{V}}_t = \mathbf{D}_t\hat{\mathbf{V}}_t{}^d\mathbf{D}_t$.

8.6 Conclusions

This chapter has outlined the multivariate temporal disaggregation methodology by which the main aggregates of the quarterly national accounts can be expressed at the monthly frequency. The challenges are mostly due to the large dimensionality of the problem. We have suggested that a methodology based upon a multivariate factor model could be capable of handling a very large number of time series subject to contemporaneous and temporal aggregation constraints.

Bibliography

Anderson, B. D.O., and Moore, J. B. (1979). Optimal Filtering. Prentice-Hall: Englewood Cliffs, NJ.

Angelini, E., Henry, J. and Marcellino, M. (2004). Interpolation and back-dating with a large information set. CEPR Discussion Paper Series 4533.

Bloem, A., Dippelsman, R.J. and Maehle, N.O. (2001), *Quarterly National Accounts Manual - Concepts, Data Sources, and Compilation*, International Monetary Fund.

Casals, J., Jerez, M. and Sotoca, S. (2005) Empirical Modeling of Time Series Sampled at Different Frequencies. European Commission (Eurostat, EuroIndicators), Working Papers and Studies.

Chow, G., and Lin, A. L. (1971). Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series, *The Review of Economics and Statistics*, 53, 4, 372-375.

Cox, D (1981). Statistical analysis of time series. Some recent developments. Scandinavian Journal of Statistics, 8, 93-115.

Dagum, E. B., Cholette, P. A. (2006). Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series. Lecture Notes in Statistics 186. Springer-Verlag New York.

Das K., Jiang J. and Rao J.N.K. (2004). Mean squared error of empirical predictor. The Annals of Statistics 2004, Vol. 32, No. 2, 818-840.

de Jong, P. (1989). Smoothing and interpolation with the state space model, *Journal of the American Statistical Association*, 84, 1085-1088.

de Jong, P. (1991). The diffuse Kalman filter, *Annals of Statistics*, 19, 1073-1083.

de Jong, P., and Chu-Chun-Lin, S. (1994). Fast Likelihood Evaluation and Prediction for Nonstationary State Space Models, *Biometrika*, 81, 133-142.

Di Fonzo, T. (1990), The estimation of M disaggregated time series when contemporaneous and temporal aggregates are known. The Review of Economics and Statistics, 72, 1, pp. 178–182

Di Fonzo, T. and Marini, M. (2005). Benchmarking a system of time series: Denton's movement preservation principle vs. a data based procedure. European Commission (Eurostat, EuroIndicators), Working Papers and Studies.

Di Fonzo, T., Marini, M. (2011), Simultaneous and Two-step Reconciliation of Systems of Time Series: Methodological and Practical Issues, *Journal of the Royal Statistical Society C*, 60, 143-164.

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, 333, pp. 99-102.

Doran, H.E. (1992). Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-varying Restrictions. Review of Economics and Statistics, 74, 568-572.

Doran, H.E. and Rambaldi, A.N. (1997). Applying Linear Time-varying Constraints to Econometric Models: with an Application to Demand Systems. Journal of Econometrics, 79, 83-95.

Durbin J. and B. Quenneville (1997), Benchmarking by state space models, International Statistical Review, 65(1): 23-48.

Eurostat (1999), Handbook of quarterly national accounts, Luxembourg, European Commission.

Fernández, P. E. B. (1981). A methodological note on the estimation of time series, *The Review of Economics and Statistics*, 63, 3, 471-478.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100, 830-40.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2011). EUROMIND: A Monthly Indicator of the Euro Area Economic Conditions, *Journal of the Royal Statistical Society - Series A*, 174, 439-470.

Grassi S., Proietti T., Frale C., Marcellino M. and Mazzi G. (2014). EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member c ountries. *International Journal of Forecasting*, accepted for publication.

Gómez, V. and Aparicio-Pérez, F. (2006). A state space approach to the problem of sampling at different frequencies. Working paper, Ministerio De Economia y Hacienda.

Jungbacker, B., and Koopman, S.J. (2012). Likelihood-based analysis for dy- 41 namic factor models. Working papers, VU University of Amsterdam. 42

Jungbacker, B., Koopman, S. J., and van der Wel, M. (2011). Dynamic factor analysis in the presence of missing data. *Journal of Economic Dynamic and Control*, 35, 1358-1368.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

Harvey, A.C. and Chung, C.H. (2000) Estimating the underlying change in unemployment in the UK. Journal of the Royal Statistics Society, Series A, 163, 303-339.

Hedhili, L., and Trabelsi, A. (2005). A polynomial method for temporal disaggregation of multivariate time series. European Commission (Eurostat, EuroIndicators), Working Papers and Studies.

Koopman, S.J., and Durbin, J. (2012). Time Series Analysis by State Space Methods: Second Edition, Oxford Statistical Science Series.

Koopman, S.J., and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models, *Journal of Time Series Analysis*, 21, 281–296.

Moauro F. and Savio G. (2005). Temporal Disaggregation Using Multivariate Structural Time Series Models. Econometrics Journal, 8, 214-234.

Palm, F. and T Nijman (1984), Missing observations in the dynamic regression model, Econometrica, 52, 1415-35.

Pandher, G.S. (2002). Forecasting Multivariate Time Series with Linear Restrictions using Constrained Structural State Space Models. Journal of Forecasting, 21, 281-300.

Pandher, G.S. (2007). Modelling and Controlling Monetary and Economic Identities with Constrained State Space Models. International Statistical Review, XX, X, 1-20

Proietti, T. (2004). Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited. Econometrics Journal, Vol. 9, pp. 357-372.

Proietti, T. (2006). On the estimation of nonlinearly aggregated mixed models. *Journal of Computational and Graphical Statistics*, Vol. 15, 1–21.

Proietti, T. (2011a). Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints, *International Statistical Review*, 79, 455–76.

Proietti, T. (2011b). Multivariate Temporal Disaggregation with Cross-Sectional constraints, *Journal of Applied Statistics*, 38, 1455–1466.

Proietti T. and Moauro F. (2006). Dynamic Factor Analysis with Nonlinear Temporal Aggregation Constraints. Journal of the Royal Statistical Society, series C (Applied Statistics), 55, pp. 281-300.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. Statistical Science. 6:15-51.4.

Rossi, N. (1982). A Note on the Estimation of Disaggregate Time Series when the Aggregate is Known, The Review of Economics and Statistics, 64, pp. 695âĂŞ6.

Stock J. and Watson M. (1989). New Indexes of Coincident and Leading Economic Indicators. In NBER Macroeconomics Annual, NBER.

Stock, J.H., and Watson M.W. (1991). A probability model of the coincident economic indicators. In *Leading Economic Indicators*, Lahiri K, Moore GH (eds), Cambridge University Press, New Yo rk.

Stock, J.H., and M.W. Watson (2002a) Macroeconomic Forecasting Using Diffusion Indexes. Journal of Business and Economic Statistics 20, pp.147-162.

Stock, J.H., and M.W. Watson (2002b) Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association 97, pp. 1167-79.

9 A Survey of Econometric Methods for Mixed-Frequency Data

Authors[†]

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

Claudia Foroni
European Central Bank
foroni.claudia@gmail.com

Table of Contents

[†]Note: The first version of this chapter was submitted in January 2013.

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

9.1 Introduction

In recent times, econometric models that take into account the information in unbalanced datasets have attracted substantial attention. Policy-makers, in particular, need to assess in real-time the current state of the economy and its expected developments, when only incomplete information is available.

In real-time, the unbalancedness of datasets arises mainly due to two features: the different sampling frequency with which the indicators are available and the so-called “ragged-edge” problem, namely, publication delays cause missing values for some of the variables at the end of the sample, see Wallis (1986). As an example, one of the key indicators of macroeconomic activity, the Gross Domestic Product (GDP), is released quarterly and with a considerable publication lag, while a range of leading and coincident indicators is available more timely and at a monthly or even higher frequency.

In this paper we review the methods proposed so far in the literature to deal with mixed-frequency data and missing values due to publication lags, see Banbura et al. (2011) and Banbura et al. (2012) for complementary overviews with a stronger focus on Kalman filter based factor modeling techniques and Harvey and Pierse (1984), Harvey (1985) and Harvey and Chung (2000) for additional details, derivations and examples.

The simplest approach is to aggregate the data to obtain a balanced dataset at the same frequency and to work with a "frozen" final vintage dataset, which eliminates the ragged edge problem. However, in the literature there are also a few methods to avoid pre-filtering associated with temporally aggregated or interpolated data, and to exploit the information contained in the large number of series available in real-time at different frequencies. In what follows, we depict the main features of the bridge models, often employed in central banks and other policy making institutions, especially for nowcasting and short-term forecasting, see e.g. Baffigi et al. (2004), Diron (2008) and Bencivelli et al. (2012). We then move to one of the main strands of the literature, mixed-data sampling (MIDAS) models, parsimonious specifications based on distributed lag polynomials, which flexibly deal with data sampled at different frequencies and provide a direct forecast of the low-frequency variable (see e.g. Ghysels et al. (2004), Clements and Galvao (2008)). Finally, we consider the state-space approaches, presenting mixed-frequency VAR (MF-VAR) and factor models. Both are system approaches that jointly describe the dynamics of the variable to be explained and of the indicators, where the use of the Kalman filter provides not only predictions of the future observations but also estimates of the current latent state (see Mariano and Murasawa (2003), Mariano and Murasawa (2010)). A natural extension in the literature is the combination of the factors with the MIDAS models, and it is based on the use of factors as explanatory variables to exploit the information in large mixed-frequency datasets. The resulting model is labelled Factor-MIDAS by Marcellino and Schumacher (2010).

For each of the alternative approaches to mixed frequency modelling listed above, we first describe their key theoretical features, and then summarize empirical applications.

This chapter is organized as follows: In section 9.2, we survey the different approaches to model mixed frequency variables. In section 9.3, we discuss the additional estimation issues arising with a ragged-edge structure of the dataset. In section 9.4 we compare the main features of the different approaches. In section 9.5 we present a summary of the most significant empirical applications in this literature. Finally, in section 9.6 we summarize and conclude.

9.2 Models for mixed-frequency data

Typical regression models relate variables sampled at the same frequency. To ensure the same frequency, researchers working with time series data either aggregate the higher-frequency observations to the lowest available frequency or interpolate the lower-frequency data to the highest available frequency, see Section 9.2.1. The most common solution in empirical applications is the former, temporal aggregation. The higher-frequency data are aggregated to the lowest-frequency by averaging or by taking a representative value (for

example, the last month of a quarter). In pre-filtering the data so that left- and right-hand variables are available at the same frequency, a lot of potentially useful information might be destroyed, and mis-specification inserted in the model. Hence, direct modelling of mixed frequency data can be useful.[1]

One of the early approaches to deal with mixed-frequency data focuses on forecasting and relies on bridge equations, see e.g. Baffigi et al. (2004), i.e. equations that link the low-frequency variables and time-aggregated indicators. Forecasts of the high-frequency indicators are provided by specific high-frequency time series models, then the forecast values are aggregated and plugged into the bridge equations to obtain the forecast of the low-frequency variable. Details are provided in section 9.2.2.

In section 9.2.3 we propose a more detailed overview of one of the most recent and competitive univariate approaches, the mixed-data sampling method originally proposed by Ghysels et al. (2004). Mixed-data sampling (MIDAS) models handle series sampled at different frequencies, where distributed lag polynomials are used to ensure parsimonious specifications. Whereas early MIDAS studies focused on financial applications, see e.g. Ghysels et al. (2006), recently this method has been employed to forecast macroeconomic time series, where typically quarterly GDP growth is forecasted by monthly macroeconomic and financial indicators, see e.g. Clements and Galvao (2008) and Clements and Galvao (2009).

Another common approach in the literature is the state-space representation of the model, where to handle data with different frequencies, the low-frequency variable is considered as a high-frequency one with missing observations. The Kalman filter and smoother is then applied to estimate the missing observations and to generate forecasts. Moreover, the dynamics of the low and high-frequency series are jointly analyzed. One of the most compelling approaches at the moment is the one proposed by Zadrozny (1988) for directly estimating a VARMA model with series sampled at different frequencies. In the same fashion, Mariano and Murasawa (2010) set what they call mixed-frequency VAR (MF-VAR from now on), i.e. they introduce a VAR model for partially latent time series and cast it in state-space form, see Section 9.2.4 for more details. Among the state-space approaches we can also list mixed-frequency factor models employed, for example, to extract an unobserved state of the economy and create a new coincident indicator or forecast and nowcast GDP, see e.g. Mariano and Murasawa (2003) and Mariano and Murasawa (2010) for small scale applications and Giannone et al. (2008) and Banbura and Rünstler (2011) for large scale models. A similar approach is also followed by Frale et al. (2010) and Frale et al. (2011): differently from the other studies, dynamic factor models are applied to a set of small datasets where variables are grouped according to economic theory and institutional considerations, rather than to the entire information set. The separate small factor models are then linked together within a state-space framework.

Finally, in section 9.2.5 we review the literature that proposes to merge the two recent strands in the mixed sampling econometrics: factor models and MIDAS approach. Marcellino and Schumacher (2010) introduce Factor-MIDAS, an approach for now- and forecasting low-frequency variables exploiting information in large sets of higher-frequency indicators.

9.2.1 Aggregation and interpolation

In most of the empirical applications, the common solution in the presence of a mixed sample frequency is to pre-filter the data so that the left- and right-hand side variables are sampled at the same frequency. In the process, a lot of potentially useful information can be destroyed and mis-specification included in the model.

The standard aggregation methods depend on the stock/flow nature of the variables and, typically, it is the average of the high-frequency variables over one low-frequency period for stocks, and the sum for flows.

[1] Wohlrabe (2009) presents another review of mixed-frequency models However, his review focuses more on the earliest attempts to tackle the mixed-frequency issues. In particular he reviews in detail the aggregation and interpolation of data, and the bridge and linkage models. In this paper, instead, we focus on the most recent developments of the literature, and especially we go through the MIDAS approach and its recent extensions, the MF-VAR in a classical and Bayesian framework, and different factor models which take into account the mixed-frequency and ragged-edge nature of the dataset.

Taking the latest available value of the higher frequency variable is another option for both stock and flow variables. The underlying assumption is that the information of the previous high-frequency periods is reflected in the latest value, representative of the whole low-frequency period.

The second option to match frequencies is the interpolation of the low frequency variables, which is rarely used. There are several different interpolating methods, see e.g. Lanning (1986), Marcellino (1998) and Angelini et al. (2006). A common approach is a two-step procedure: first missing data are interpolated, then model parameters are estimated using the new augmented series, possibly taking into account the measurement error induced by disaggregation. Both steps can be conveniently and jointly run in a Kalman filter set-up, starting with a state-space representation of the model, see e.g. Harvey (1989) and Sections 9.2.4 and 9.2.5 below.

9.2.2 Bridge equations

One of the early econometric approaches in the presence of mixed-frequency data relies on the use of bridge equations, see e.g. Baffigi et al. (2004) and Diron (2008). Bridge equations are linear regressions that link ("bridge") high frequency variables, such as industrial production or retail sales, to low frequency ones, e.g. the quarterly real GDP growth, providing some estimates of current and short-term developments in advance of the release. The "Bridge model" technique allows computing early estimates of the low-frequency variables by using high frequency indicators. They are not standard macroeconometric models, since the inclusion of specific indicators is not based on causal relations, but on the statistical fact that they contain timely updated information. In principle, bridge models require that the whole set of regressors should be known over the projection period, allowing for an estimate only of the current period. In practice, anyway, this is not the case, even though the forecasting horizon of the bridge models is quite short, one or two quarters ahead at most.

Taking forecasting GDP as an example, since the monthly indicators are usually only partially available over the projection period, the predictions of quarterly GDP growth are obtained in two steps. First, monthly indicators are forecasted over the remainder of the quarter, usually on the basis of univariate time series models (in some cases VAR have been implemented in order to obtain better forecasts of the monthly indicators), and then aggregated to obtain their quarterly correspondent values. Second, the aggregated values are used as regressors in the bridge equation which allows to obtain forecasts of GDP growth.

Therefore, the bridge model to be estimated is:

$$y_{t_q} = \alpha + \sum_{i=1}^{j} \beta_i (L) x_{it_q} + u_{t_q} \tag{9.1}$$

where $\beta_i (L)$ is a lag polynomial of length k, and x_{it_q} are the selected monthly indicators aggregated at quarterly frequency.

The selection of the monthly indicators included in the bridge model is usually based on a general-to-specific methodology and relies on different in-sample or out-of-sample criteria, like information criteria or RMSE performance. Bencivelli et al. (2012) propose an alternative procedure based on Bayesian Model Averaging (BMA) that performs quite well empirically.

In order to forecast the missing observations of the monthly indicators which are then aggregated to obtain a quarterly value of x_{it_q}, it is common practice to use autoregressive models, where the lag length is based on information criteria.

9.2.3 Mixed-Data Sampling

Distributed lag (DL) models have been typically employed in the literature to describe the distribution over time of the lagged effects of a change in the explanatory variable. In general, a stylized distributed lag model is

given by

$$y_{t_q} = \alpha + B(L)\, x_{t_q} + \varepsilon_{t_q} \tag{9.2}$$

where $B(L)$ is some finite or infinite lag polynomial operator.

This kind of models underlies the construction of the bridge equations, once all the high frequency values are aggregated to the corresponding low-frequency values.

In order to take into account mixed-frequency data, Ghysels et al. (2004) introduce the Mixed-Data Sampling (MIDAS) approach, which is closely related to the distributed lag model, but in this case the dependent variable y_{t_q}, sampled at a lower-frequency, is regressed on a distributed lag of x_{t_m}, which is sampled at a higher-frequency.

In what follows, we first present the basic features of the model as presented by Ghysels et al. (2004), the corresponding unrestricted version as in Foroni et al. (2014), and then the extensions that have been introduced in the literature.

In terms of notation, $t_q = 1, ...T_q$ indexes the basic time unit (e.g. quarters), and m is the number of times the higher sampling frequency appears in the same basic time unit. For example, for quarterly GDP growth and monthly indicators as explanatory variables, $m = 3$. w is the number of monthly values of the indicators that are earlier available than the lower-frequency variable to be estimated. The lower-frequency variable can be expressed at the high frequency by setting $y_{t_m} = y_{t_q}, \forall t_m = m t_q$, where t_m is the time index at the high frequency.

The basic MIDAS model

MIDAS regressions are essentially tightly parameterized, reduced form regressions that involve processes sampled at different frequencies. The response to the higher-frequency explanatory variable is modelled using highly parsimonious distributed lag polynomials, to prevent the proliferation of parameters that might otherwise result, as well as the issues related to lag-order selection.

The basic MIDAS model for a single explanatory variable, and h_q-step-ahead forecasting, with $h_q = h_m/m$, is given by:

$$y_{t_q + m h_q} = y_{t_m + h_m} = \beta_0 + \beta_1 b(L_m; \theta)\, x^{(m)}_{t_m + w} + \varepsilon_{t_m + h_m} \tag{9.3}$$

where $b\left(L^{1/m}; \theta\right) = \sum_{k=0}^{K} c(k; \theta)\, L_m^k$, and $L_m^x x_{t_m}^{(m)} = x_{t_m - x}^{(m)}$. $x_{t_m + w}^{(m)}$ is skip-sampled from the high frequency indicator x_{t_m}.

The parameterization of the lagged coefficients of $c(k; \theta)$ in a parsimonious way is one of the key MIDAS features. One of the most used parameterizations is the one known as "Exponential Almon Lag" , since it is closely related to the smooth polynomial Almon lag functions that are used to reduce multicollinearity in the Distributed Lag literature. It is often expressed as

$$c(k; \theta) = \frac{\exp\left(\theta_1 k + ... + \theta_Q k^Q\right)}{\sum_{k=1}^{K} \exp\left(\theta_1 k + ... + \theta_Q k^Q\right)} \tag{9.4}$$

This function is known to be quite flexible and can take various shapes with only a few parameters. These include decreasing, increasing or hump-shaped patterns. Ghysels et al. (2005) use the functional form with two parameters, which allows a great flexibility and determines how many lags are included in the regression.

Notice that the standard practice in bridge equations of calculating a quarterly series from the monthly indicators corresponds to imposing restrictions on this parameterization function. To be concrete, in the case of the quarterly-monthly example, taking the last month in the quarter to produce a quarterly series amounts to setting $c(2; \theta) = c(3; \theta) = c(5; \theta) = c(6; \theta) = ... = c(11; \theta) = c(12; \theta) = 0$.

Another possible parameterization, also with only two parameters, is the so-called "Beta Lag" , because it is based on the Beta function:

$$c\left(k;\theta_1,\theta_2\right)=\frac{f\left(\frac{k}{K},\theta_1;\theta_2\right)}{\sum_{k=1}^{K}f\left(\frac{k}{K},\theta_1;\theta_2\right)} \tag{9.5}$$

where $c\left(x,a,b\right)=\frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ and $\Gamma\left(a\right)=\int_0^\infty e^{-x}x^{a-1}dx$.

Ghysels et al. (2009) propose also three other different parameterizations of the lag coefficients: a linear scheme, with $c\left(k;\theta\right)=\frac{1}{K}$, where there are no parameters to estimate in the lagged weight function; an hyperbolic scheme, with $c\left(k;\theta\right)=\frac{g\left(\frac{k}{K},\theta\right)}{\sum_{k=1}^{K}g\left(\frac{k}{K},\theta\right)}$, $g\left(k,\theta\right)=\frac{\Gamma(k+\theta)}{\Gamma(k+1)\Gamma(\theta)}$ where the gamma function has only one parameter to estimate, but it's not as flexible as the Beta specification; a geometric scheme, with $c\left(k;\theta\right)=\frac{\theta^k}{\sum_{k=1}^{\infty}\theta^k}$, $|\theta|\leq 1$ and $c\left(k;\theta\right)$ are normalized so that they sum up to one.

The parameterizations described above are all quite flexible. For different values of the parameters, they can take various shapes: weights attached to the different lags can decline slowly or fast, or even have a hump shape. Therefore, estimating the parameters from the data automatically determines the shape of the weights and, accordingly, the number of lags to be included in the regression.

The MIDAS model can be estimated using nonlinear least squares (NLS) in a regression of y_t onto $x_{t-h}^{(m)}$. Ghysels et al. (2004) show that MIDAS regressions always lead to more efficient estimation than the typical approach of aggregating all series to the least frequent sampling. Moreover, they also show that discretization biases are the same for MIDAS and distributed lag models and vanish when regressors are sampled more frequently.

The forecast is given by

$$\widehat{y}_{T_m^y+h_m|T_m^x}=\widehat{\beta}_0+\widehat{\beta}_1 b\left(L_m;\widehat{\theta}\right)x_{T_m^x}^{(m)}. \tag{9.6}$$

Note that MIDAS is $h-$dependent, and thus needs to be re-estimated for each forecast horizon.

The AR-MIDAS model

Since autoregressive models often provide competitive forecasts to those obtained with models that include explanatory variables, the introduction of an autoregressive term in the MIDAS model is a desirable extension, although not straightforward. Ghysels et al. (2004) show that the introduction of lagged dependent variables creates efficiency losses. Moreover, it would result in the creation of seasonal patterns in the explanatory variables.

Consider adding a lower-frequency lag of y_{t_m}, y_{t_m-3}, to the basic model with $m=3$ (x is monthly and y is quarterly):

$$y_{t_m}=\beta_0+\lambda y_{t_m-3}+\beta_1 b\left(L_m;\theta\right)x_{t_m+w-3}^{(3)}+\varepsilon_{t_m}. \tag{9.7}$$

As highlighted in Clements and Galvao (2009), this strategy is in general not appropriate. The reason becomes clear when we write the model as:

$$y_{t_m}=\beta_0\left(1-\lambda\right)^{-1}+\beta_1\left(1-\lambda L_m^3\right)^{-1}B\left(L_m;\theta\right)x_{t_m+w-3}^{(3)}+\left(1-\lambda L_m^3\right)^{-1}\varepsilon_{t_m}. \tag{9.8}$$

The polynomial on $x_{t-1}^{(3)}$ is a product of a polynomial in $L^{1/3}$ and a polynomial in L. This product generates a seasonal response of y to $x^{(3)}$, irrespective of whether $x^{(3)}$ displays a seasonal pattern.

To avoid this inconvenience, the authors suggest the introduction of the AR dynamics as a common factor:

$$y_{t_m} = \beta_0 + \lambda y_{t_m-3} + \beta_1 b\left(L_m;\theta\right)\left(1-\lambda L_m^3\right) x_{t_m+w-3}^{(3)} + \varepsilon_{t_m} \tag{9.9}$$

so that the response of y to $x^{(3)}$ remains non-seasonal.

The analogous multi-step model is written as:

$$y_{t_m} = \beta_0 + \lambda y_{t_m-h_m} + \beta_1 b\left(L_m;\theta\right)\left(1-\lambda L_m^{h_m}\right) x_{t+w-h_m}^{(3)} + \varepsilon_{t_m}. \tag{9.10}$$

To estimate the MIDAS-AR model, the common procedure is:

- To estimate the standard MIDAS (the basic model), take the residuals $\widehat{\varepsilon}_{t_m}$;
- To estimate an initial value for λ, say λ_0, where $\widehat{\lambda}_0 = \left(\sum \widehat{\varepsilon}_{t_m+w-h_m}^2\right)^{-1} \sum \widehat{\varepsilon}_{t_m}\widehat{\varepsilon}_{t_m+w-h_m}$.
- Then construct $y_{t_m}^* = y_{t_m} - \widehat{\lambda}_0 y_{t_m-h_m}$ and $x_{t_m+w-h_m}^{*(3)} = x_{t_m+w-h_m}^{(3)} - \widehat{\lambda}_0 x_{t_m-(h_m-w)-h_m}^{(3)}$.
- The estimator $\widehat{\theta}_1$ is obtained by applying nonlinear least squares to:

$$y_{t_m}^* = \beta_0 + \beta_1 b\left(L_m;\theta\right) x_{t_m+w-h_m}^{*(3)} + \varepsilon_{t_m}. \tag{9.11}$$

- A new value of λ, $\widehat{\lambda}_1$, is obtained from the residuals of this regression.
- Then a new step is run, using $\widehat{\lambda}_1$ and $\widehat{\theta}_1$ as the initial values.

In this way, the procedure gets the estimates and $\widehat{\lambda}$ and $\widehat{\theta}$ that minimize the sum of squared residuals.

The Unrestricted MIDAS model

Foroni et al. (2014) study the performance of a variant of MIDAS which does not resort to functional distributed lag polynomials. In the paper, the authors discuss how unrestricted MIDAS (U-MIDAS) regressions can be derived in a general linear dynamic framework, and under which conditions the parameters of the underlying high-frequency model can be identified[2].

The U-MIDAS model based on a linear lag polynomial such as

$$\begin{aligned} c(L^m)\omega(L)y_{t_m} &= \delta_1(L)x_{1t_m-1} + ... + \delta_N(L)x_{Nt_m-1} + \epsilon_{t_m}, \\ t &= 1,2,3,... \end{aligned} \tag{9.12}$$

where $c(L^m) = (1 - c_1L^m - ... - c_cL^{mc})$, $\delta_j(L) = (\delta_{j,0} + \delta_{j,1}L + ... + \delta_{j,v}L^v)$, $j = 1, ..., N$.

Note that if we assume that the lag orders c and v are large enough to make the error term ϵ_{t_m} uncorrelated, then, all the parameters in the U-MIDAS model (9.12) can be estimated by simple OLS (while the aggregation scheme $\omega(L)$ is supposed known). From a practical point of view, the lag order v could differ across variables, and v_i and c could be selected by an information criterion such as BIC.

A simple approach to forecasting is to use a form of direct estimation and construct the forecast as

$$\widetilde{y}_{T_m^x+m|T_m^x} = \widetilde{c}(L^k)y_{T_m^x} + \widetilde{\delta}_1(L)x_{1T_m^x} + ... + \widetilde{\delta}_N(L)x_{NT_m^x}, \tag{9.13}$$

where the polynomials $\widetilde{c}(Z) = \widetilde{c}_1L^m + ... + \widetilde{c}_cL^{mc}$ and $\widetilde{\delta}_i(L)$ are obtained by projecting y_{t_m} on information dated $mt_m - m$ or earlier, for $t = 1, 2, ..., T_m^x$. In general, the direct approach of (9.13) can also be extended

[2] Koenig et al. (2003) already proposed U-MIDAS in the context of real-time estimation. However, they did not systematically study the role of the functional form of the lag polynomial.

to construct h_m-step ahead forecasts given information in T_m^x:

$$\overline{y}_{T_m^x+h_m|T_m^x} = \overline{c}(L^k)y_{T_m^x} + \overline{\delta}_1(L)x_{1T_m^x} + ... + \overline{\delta}_N(L)x_{NT_m^x}, \tag{9.14}$$

where the polynomials $\overline{c}(Z)$ and $\overline{\delta}_i(L)$ are obtained by projecting y_{t_m} on information dated $mt-h_m$ or earlier, for $t = 1, 2, ..., T_m^x$.

In the case of U-MIDAS, the autoregressive term can be included easily without any common factor restriction as in Clements and Galvao (2009).

Finally, Carriero et al. (2013) use Bayesian techniques to estimate specifications similar to U-MIDAS models with several regressors and stochastic volatility, which can easily produce not only point but also interval and density forecasts. We refer to their paper for the technical details.

Extensions of the MIDAS model

Different extensions of the MIDAS models have been analyzed in the literature, to introduce the use of mixed-frequency data in specific applications or studies, in which there is a need to capture particular features. For example, some studies incorporate regime changes in the parameters or asymmetric reactions to negative or positive values of the explanatory variables.

In what follows, we provide a brief overview of the extensions of the MIDAS models discussed so far in the literature.

Multiple explanatory variables

To allow for the inclusion of several additional explanatory variables into the MIDAS framework, it is necessary to extend the basic model above as follows:

$$y_{t_m} = \beta_0 + \beta_1 b\left(L_m;\theta_1\right) x_{1,t_m+w-h_m}^{(m)} + \beta_2 b\left(L_m;\theta_2\right) x_{2,t_m+w-h_m}^{(m)} + \varepsilon_{t_m}. \tag{9.15}$$

In this case, we consider x_1 and x_2 as two different explanatory variables. The values of the theta parameters are assumed to take on independent values and are thus represented by two independent vectors for the parameters, which may have different lag lengths.

Obviously, the above specification may be extended to allow for the inclusion of more than two explanatory variables (or more than two lags), and for the presence of an autoregressive structure. The most general MIDAS linear regression model can then be written as

$$y_{t_m} = \beta_0 + \sum_{i=1}^{K}\sum_{j=1}^{L} b_{ij}\left(L_{m_i};\theta\right) x_{t_m+w-h_m}^{(m_i)} + \varepsilon_{t_m}. \tag{9.16}$$

Within the more general framework, it is also possible to include explanatory variables at different frequencies, since each indicator is modelled with its own polynomial parameterization. As an example, quarterly GDP growth can be explained not only by monthly indicators but also by weekly financial variables, with the explanatory variables, therefore, sampled at two different frequencies.

Nonlinear MIDAS models

Ghysels et al. (2007) further generalize (9.16) to:

$$y_{t_m} = \beta_0 + f\left(\sum_{i=1}^{K}\sum_{j=1}^{L} b_{ij}\left(L_{m_i};\theta\right) g\left(x_{t_m+w-h_m}^{(m_i)}\right)\right) + \varepsilon_{t_m}, \tag{9.17}$$

where the functions f and g can be either fully known or parameter dependent. This model is inspired by the EGARCH model, and can be useful especially in volatility applications and risk-return trade-off studies.

Asymmetric MIDAS models

Ghysels et al. (2005) introduce the asymmetric MIDAS model given by:

$$y_{t_m} = \beta_0 + \beta_1 \left(\phi b \left(L_m; \theta^- \right) 1^-_{t_m - h_m} x^{(m)}_{t_m + w - h_m} + (2 - \phi)\, b \left(L_m; \theta^+ \right) 1^+_{t_m - h_m} x^{(m)}_{t_m + w - h_m} \right) + \varepsilon_{t_m} \quad (9.18)$$

where $1^+_{t_m - h_m}$ denotes the indicator function for $x^{(m)}_{t_m + w - h_m} \geq 0$ and $1^-_{t_m - h_m}$ for $x^{(m)}_{t_m + w - h_m} < 0$, and $\phi \in (0, 2)$ in order to ensure that the total weights sum up to one. This formulation allows for a different impact of negative and positive values of the regressor x. The value of ϕ controls the different weight put on positive and negative impacts. Allowing for an asymmetric impact of the indicator is important in financial applications, especially in examining the asymmetric reaction of volatility in positive and negative return shocks.

Smooth Transition MIDAS models

Galvao (2007) proposes a new regression model which combines a smooth transition regression with a mixed data sampling approach:

$$\begin{aligned} y_{t_m} \quad = \quad & \beta^{(m)}_{0,h_m} + \beta^{(m)}_{1,h_m} x^{(m)}_{t_m + w - h_m} \left[1 - G_{t_m + w - h_m} \left(x^{(m)}_{t_m + w - h_m}; \gamma, c \right) \right] + \\ & + \beta^{(m)}_{2,h_m} x^{(m)}_{t_m + w - h_m} \left[G_{t_m + w - h_m} \left(x^{(m)}_{t_m + w - h_m}; \gamma, c \right) \right] + \varepsilon_{t_m} \end{aligned} \quad (9.19)$$

where

$$G_{t_m + w - h_m} \left(x^{(m)}_{t_m}; \gamma, c \right) = \frac{1}{1 + \exp \left(-\gamma / \widehat{\sigma}_x \left(x^{(m)}_{t_m + w - h_m}; \gamma, c \right) \right)} \quad (9.20)$$

The transition function is a logistic function that depends on the weighted sum of the explanatory variable in the current quarter.

The time-varying structure allows for changes in the predictive power of the indicators. This can be particularly relevant when one wants to use asset returns for forecasting macroeconomic variables, since changes in the predictive power of asset returns on economic activity may be related to business cycle regimes.

Markov-Switching MIDAS models

Guerin and Marcellino (2013) incorporate regime changes in the parameters of the MIDAS models. The basic version of the Markov-Switching MIDAS (MS-MIDAS) regression model they propose is:

$$y_{t_m} = \beta_0 \left(S_{t_m} \right) + \beta_1 \left(S_{t_m} \right) B \left(L_m; \theta \right) x^{(m)}_{t_m + w - h_m} + \varepsilon_{t_m} \left(S_{t_m} \right) \quad (9.21)$$

where $\varepsilon_{t_m} | S_{t_m} \sim NID \left(0, \sigma^2 \left(S_{t_m} \right) \right)$. The regime generating process is an ergodic Markov-chain with a finite number of states S_{t_m}.

These models allow also mixed-sample estimation of the probabilities of being in a given regime, which are relevant, for example, when one wants to predict business cycle regimes. The model is extended to the VAR context by Foroni et al. (2014).

MIDAS with step functions

Forsberg and Ghysels (2007) introduce a MIDAS regression with step functions, where the distributed lag pattern is approximated by a number of discrete steps. To define this MIDAS regression, we consider the regressors $X \left(t_m, K_i \right) = \sum_{j=1}^{K_i} x^{(m)}_{t_m - j}$, which are partial sums of the high frequency variables. Then the MIDAS

regression with M steps is:

$$y_{t_m} = \beta_0 + \sum_{i=1}^{M} \beta_i X\left(t_m, K_i\right) + \varepsilon_{t_m}. \tag{9.22}$$

This special case of MIDAS models can be reconnected to the U-MIDAS case we have analyzed in Section 9.2.3, in which the steps are the single individual observations.

Multivariate MIDAS models

Regression (9.16) can be generalized to multivariate specifications:

$$Y_{t_m} = \mathcal{B}_0 + \sum_{i=1}^{K}\sum_{j=1}^{L} \mathcal{B}_{ij}\left(L_{m_i}; \theta\right) X_{t_m+w-h_m}^{(m_i)} + \varepsilon_{t_m}, \tag{9.23}$$

where Y, ε and X are $n-$dimensional vector processes B_0 is an $n-$dimensional vector and B_{ij} are $n \times n$ matrices of polynomials. The main issue is how to handle parameter proliferation in a multivariate context. One approach is to consider all the off-diagonal elements controlled by one polynomial, while the diagonal elements by a second one. Of course, the restrictions may not be valid, and will be chosen depending on the application.

Considering multivariate MIDAS regressions allows to address Granger causality issues, avoiding temporal aggregation errors that can disguise or create spurious causality.

9.2.4 Mixed-frequency VAR

While so far, we have seen models which take into account mixed-frequency data in a univariate approach, we now focus on multivariate methods which jointly specify the dynamics of the indicators and of the variable to be explained. To exploit the information available in series released at different frequencies and jointly analyze them, there is a growing literature which looks at mixed-frequency VARs, which aim to characterize the co-movements in the series and summarize the information contained in the mixed-frequency data.

Nowadays, in the literature, there are both classical and Bayesian approaches to estimate MF-VAR models. In what follows, we describe the main features of these two classes of estimation, following two of the most representative studies in the literature, Mariano and Murasawa (2010) for the classical approach and Schorfheide and Song (2011) for the Bayesian approach.

Classical framework

One of the most compelling approaches in the literature to deal with mixed-frequency time series at the moment is the one proposed by Zadrozny (1988) for directly estimating a VARMA model sampled at different frequencies, see also Harvey (1989). The approach treats all the series as generated at the highest frequency, but some of them are not observed. Those variables that are observed only at the low frequency are therefore considered as periodically missing.

Following the notation of Mariano and Murasawa (2010), we consider the state-space representation of a VAR model in a classical framework, treating quarterly series as monthly series with missing observations and taking GDP growth as an example. The disaggregation of the quarterly GDP growth, y_{t_m}, observed every $t_m = 3, 6, 9, ..., T_m$, into the month-on-month GDP growth, $y_{t_m}^*$, never observed, is based on the following aggregation equation:

$$\begin{aligned} y_{t_m} &= \frac{1}{3}\left(y^*_{t_m} + y^*_{t_m-1} + y^*_{t_m-2}\right) + \frac{1}{3}\left(y^*_{t_m-1} + y^*_{t_m-2} + y^*_{t_m-3}\right) + \\ &\quad + \frac{1}{3}\left(y^*_{t_m-2} + y^*_{t_m-3} + y^*_{t_m-4}\right) \\ &= \frac{1}{3}y^*_{t_m} + \frac{2}{3}y^*_{t_m-1} + y^*_{t_m-2} + \frac{2}{3}y^*_{t_m-3} + \frac{1}{3}y^*_{t_m-4}. \end{aligned} \tag{9.24}$$

This aggregation equation comes from the assumption that the quarterly GDP series (in log levels), Y_{t_m}, is the geometric mean of the latent monthly random sequence $Y^*_{t_m}, Y^*_{t_m-1}, Y^*_{t_m-2}$. Taking the three-period differences and defining $y_{t_m} = \Delta_3 Y_{t_m}$ and $y^*_{t_m} = \Delta Y^*_{t_m}$, we obtain eq. (9.24).

Let for all t_m the latent month-on-month GDP growth $y^*_{t_m}$ and the corresponding monthly indicator x_{t_m} follow a bivariate VAR(p) process

$$\phi\left(L_m\right)\begin{pmatrix} y^*_{t_m} - \mu^*_y \\ x_{t_m} - \mu_x \end{pmatrix} = u_{t_m}, \tag{9.25}$$

where $u_{t_m} \sim N\left(0, \Sigma\right)$.

The VAR(p) process in eq. (9.25) together with the aggregation equation (9.24) is then cast in a state-space representation.

Assuming $p \leq 4$[3] and defining

$$s_{t_m} = \begin{pmatrix} z_{t_m} \\ \vdots \\ z_{t_m-4} \end{pmatrix}, \quad z_{t_m} = \begin{pmatrix} y^*_{t_m} - \mu^*_y \\ x_{t_m} - \mu_x \end{pmatrix},$$

a state-space representation of the MF-VAR is

$$s_{t_m} = Fs_{t_m-1} + Gv_{t_m} \tag{9.26}$$

$$\begin{pmatrix} y_{t_m} - \mu_y \\ x_{t_m} - \mu_x \end{pmatrix} = Hs_{t_m} \tag{9.27}$$

with $\mu_y = 3\mu^*_y$ that holds, and $v_{t_m} \sim N\left(0, I_2\right)$ The matrices are defined as:

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}; \quad F_1 = \begin{bmatrix} \phi_1 & \ldots & \phi_p & 0_{2\times 2(5-p)} \end{bmatrix}; \quad F_2 = \begin{bmatrix} I_8 & 0_{8\times 2} \end{bmatrix}, \tag{9.28}$$

$$G = \begin{bmatrix} \Sigma^{1/2} \\ 0_{8\times 2} \end{bmatrix}; \quad H = \begin{bmatrix} H_0 & \ldots & H_4 \end{bmatrix} \tag{9.29}$$

where H contains the lag polynomial

$$H\left(L_m\right) = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L_m + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} L^2_m + \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L^3_m + \begin{bmatrix} 1/3 & 0 \\ 0 & 0 \end{bmatrix} L^4_m \tag{9.30}$$

The state-space model consisting of equations (9.26) and (9.27) can be estimated with maximum-likelihood techniques or the expectation-maximization algorithm, where we have to take into account missing values

[3] For the sake of conciseness, we do not report the state-space representation for $p > 4$. Details for this case can be found in Mariano and Murasawa (2010).

due to publication lags and the low-frequency nature of the GDP. We illustrate the estimation and forecasting issues later on, in section 9.3, where we review the problems related to ragged-edge data.

Bayesian framework

The estimation of MF-VAR model with Bayesian techniques has been recently considered as an alternative framework in the literature. One of the earliest studies on this is the paper by Chiu et al. (2011). In this paper, the authors develop a Gibbs sampling approach to estimate a VAR with mixed and irregularly sampled data. The algorithm they develop is a Gibbs sampler which iterates over the draws from the missing data and from the unknown parameters in the model. Under the assumption of a normally distributed error term, the algorithm allows for draws from Gaussian conditional distributions for estimating the missing data, and for draws from Gaussian and inverse Wishart conditional posterior distributions for the parameters in the model.

As an example for the Bayesian estimation of a MF-VAR, we present the algorithm developed by Schorfheide and Song (2011). The authors represent the MF-VAR as a state-space model, and use MCMC methods to conduct Bayesian inference for model parameters and unobserved monthly variables.

The state equation of the model is represented by the VAR(p) model written in the companion form:

$$z_{t_m} = F_1(\Phi) z_{t_m-1} + F_c(\Phi) + v_{t_m}, \quad v_{t_m} \sim iidN(0, \Omega(\Sigma)). \tag{9.31}$$

To write the measurement equation, the authors need to write the aggregation equation, which is in this case different from the one considered by Mariano and Murasawa (2010). In this case, the quarterly variable is seen as the three-month average of the monthly process, which in the previous notation is:

$$y_{t_m} = \frac{1}{3}\left(y^*_{t_m} + y^*_{t_m-1} + y^*_{t_m-2}\right) = \Lambda_{mz} z_{t_m}. \tag{9.32}$$

However, since y_{t_m} is observed only every third month, there is a need of a selection matrix that equals the identity matrix if t_m corresponds to the last month of the quarter and is empty otherwise. Therefore, the measurement equation can be written as

$$\begin{pmatrix} y_{t_m} \\ x_{t_m} \end{pmatrix} = M_{t_m} \Lambda_z z_{t_m}, \tag{9.33}$$

where M_{t_m} is the selection matrix. A Minnesota prior that shrinks the VAR coefficients toward univariate random walk representations is introduced to cope with the issue of dimensionality.

9.2.5 Mixed-frequency factor models

Closely related to the MF-VAR for their state-space representation, factor models have also been employed in the literature to handle data with different frequencies. These models have been utilized to extract an unobserved state of the economy and create a new coincident indicator, but also to exploit more information and obtain more precise forecasts. In what follows, we discuss the Mariano and Murasawa (2003) small scale mixed-frequency factor model, developed to extend the Stock–Watson coincident index for the US economy by combining quarterly real GDP and monthly coincident business cycle indicators. Interesting applications of a similar approach can be found in Frale et al. (2010) and Frale et al. (2011). Then, we present an example of large scale mixed-frequency factor model, as proposed by Giannone et al. (2008), whose aim is to bridge the information in a large monthly dataset with the forecast of a quarterly variable. As an extension to it, we present the mixed-frequency state-space framework as developed by Banbura and Rünstler (2011). Finally, based on Marcellino and Schumacher (2010), we analyze the approach that merges factor models and the MIDAS framework presented above.

Mixed-frequency small scale factor models

Factor models have a long tradition in econometrics and they are also appealing from an economic point of view. In fact, they decompose each time series under analysis into a common component, driven by few factors that represent the key economic driving forces, and an idiosyncratic component.

Mariano and Murasawa (2003) set up a static one-factor model for a small set of observable monthly and quarterly series, and derive its state-space representation.

Following their notation, consider a one-factor model for y_t^*, such that for all t_m,

$$\begin{aligned} y_{t_m}^* &= \mu^* + \Lambda f_{t_m} + u_{t_m} && (9.34)\\ \Phi_f(L) f_{t_m} &= v_{t_m} && (9.35)\\ \Phi_u(L) u_{t_m} &= w_{t_m} && (9.36)\\ \begin{pmatrix} v_{t_m} \\ w_{t_m} \end{pmatrix} &\sim N\left(0, \begin{bmatrix} \Sigma_{vv} & 0 \\ 0 & \Sigma_{ww} \end{bmatrix}\right) && (9.37) \end{aligned}$$

where $\Phi_f(.)$ is a pth-order polynomial on $\mathbb{R}$ and $\Phi_u(.)$ is a qth-order polynomial on $\mathbb{R}^{N\times N}$. In order to have identification, we assume $\Lambda := [I, \Lambda_2']'$ and $\Phi_u(.)$ and Σ_{ww} diagonal.

State-space representation

Assuming p,$q \leq 4$, for all t_m, and defining

$$s_t = \begin{pmatrix} f_{t_m} \\ \vdots \\ f_{t_m-4} \\ u_{t_m} \\ \vdots \\ u_{t_m-4} \end{pmatrix},$$

the state-space representation of the factor model is

$$\begin{aligned} s_{t_m+1} &= F s_{t_m} + G v_{t_m} && (9.38)\\ y_{t_m} &= \mu + H s_{t_m} && (9.39) \end{aligned}$$

with $v_{t_m} \sim N(0, I_3)$, where

$$\begin{aligned} F &= \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix};\ F_1 = \begin{bmatrix} \Phi_{f,1}...\Phi_{f,p} & 0_{1\times(5-p)} \\ I_4 & 0_{4\times1} \end{bmatrix};\ F_2 = 0_{5\times10}; && (9.40)\\ F_3 &= 0_{10\times5};\ F_4 = \begin{bmatrix} \Phi_{u,1}...\Phi_{u,q} & 0_{1\times(5-q)} \\ I_8 & 0_{8\times2} \end{bmatrix} \\ G &= \begin{bmatrix} \Sigma_{vv}^{1/2} & 0_{1\times2} \\ 0_{4\times1} & 0_{4\times2} \\ 0_{2\times1} & \Sigma_{ww}^{1/2} \\ 0_{8\times1} & 0_{8\times2} \end{bmatrix};\ H = \begin{bmatrix} H_0\Lambda & ... & H_4\Lambda & H_0 & ... & H_4 \end{bmatrix} && (9.41) \end{aligned}$$

where $H(L_m)$ is defined as in equation (9.30).

In the estimation, Mariano and Murasawa (2003) cannot use the standard EM algorithm, since the measurement equation has unknown parameters. The procedure they followed is similar to the one described in

Section 9.3.1.

The dynamic factor model as extended by Mariano and Murasawa (2003) is also used in Frale et al. (2011) to handle mixed frequency data, in order to obtain estimates of the monthly Euro area GDP components from the output and expenditure sides, to be later aggregated into a single indicator, called EUROMIND. Broadly speaking, GDP is disaggregated by supply sectors and demand components. For each of these sectors and components, timely and economically sensible observable monthly indicators are then selected and represented with a dynamic factor model, as described above. The single models are then linked together based on the composition of GDP.

Bridging with factors

We now discuss a large mixed frequency factor model as proposed by Giannone et al. (2008), which exploits a large number of series that are released at different times and with different lags. The methodology the authors propose relies on the two-step estimator by Doz et al. (2011). This framework combines principal components with the Kalman filter. First, the parameters of the model are estimated by OLS regression on the estimated factors, where the latter are obtained through principal components calculated on a balanced version of the dataset. Then, the Kalman smoother is used to update the estimate of the signal variable on the basis of the entire unbalanced panel.

The model

The dynamic factor model of Doz et al. (2011) is given by

$$x_{t_m} = \Lambda f_{t_m} + \xi_{t_m} \quad \xi_{t_m} \sim N(0, \Sigma_\xi) \tag{9.42}$$

$$f_{t_m} = \sum_{i=1}^{p} A_i f_{t_m - i} + B\eta_{t_m} \quad \eta_{t_m} \sim N(0, I_q) \tag{9.43}$$

Equation (9.42) relates the N monthly series x_{t_m} to a $r \times 1$ vector of latent factors f_{t_m}, through a matrix of factor loadings Λ, plus an idiosyncratic component ξ_{t_m}, assumed to be a multivariate white noise with diagonal covariance matrix Σ_ξ. Equation (9.43) describes the law of motion of the latent factors, which are driven by a $q-$dimensional standardized white noise η_{t_m}, where B is a $r \times q$ matrix ($r \leq q$). Hence, $\zeta_{t_m} \sim N(0, BB')$.

To deal with missing observations at the end of the sample, the authors use a two-step estimator. In the first step, the parameters of the model are estimated consistently through principal components on a balanced panel, created by truncating the data set at the date of the least timely release. In the second step, the Kalman smoother is applied to update the estimates of the factor and the forecast on the basis of the entire unbalanced data set (see section 9.4 for more details on the estimation method).

The model is then complemented by a forecast equation for mean-adjusted quarterly GDP. The forecast is defined as the projection of the quarterly GDP growth on the quarterly aggregated estimated common factors:

$$\widehat{y}_{t_q} = \alpha + \beta \widehat{f}_{t_q}, \tag{9.44}$$

where $\widehat{f}_{t_q}$ is the quarterly aggregated correspondent of $\widehat{f}_{t_m}$.

If we look at eq. (9.44), we see that this is exactly what we analyzed in Section (9.2.2) for the bridge equations. The framework can be interpreted as a large bridge model which uses a large number of variables and bridges monthly data releases with the forecast of the quarterly variable.

Factor models in a mixed-frequency state-space representation

Banbura and Rünstler (2011) extend the model of Giannone et al. (2008), by integrating a forecast equation for quarterly GDP. More specifically, they introduce the forecast of monthly GDP growth y_{t_m} as a latent variable,

related to the common factors by the static equation

$$y_{t_m} = \beta' f_{t_m} + \varepsilon_{t_m}, \quad \varepsilon_{t_m} \sim N\left(0, \sigma_\varepsilon^2\right). \tag{9.45}$$

The quarterly GDP growth, y_{t_m}, is assumed to be the quarterly average of the monthly series:

$$y_{t_m} = \frac{1}{3}\left(y^*_{t_m} + y^*_{t_m-1} + y^*_{t_m-2}\right). \tag{9.46}$$

The innovations $\varepsilon_{t_m}, \eta_{t_m}, \xi_{t_m}$ are assumed to be mutually independent at all leads and lags.

Equations (9.42) to (9.46) can be cast in state-space form. y_{t_m} is constructed in such a way that it contains the quarterly GDP growth in the third month of each quarter, while the other observations are treated as missing.

State-space representation

The state-space representation, when $p = 1$, is:

$$\begin{bmatrix} x_{t_m} \\ y_{t_m} \end{bmatrix} = \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_{t_m} \\ y^*_{t_m} \\ y^C_{t_m} \end{bmatrix} + \begin{bmatrix} \xi_{t_m} \\ \varepsilon_{t_m} \end{bmatrix} \tag{9.47}$$

$$\begin{bmatrix} I_r & 0 & 0 \\ -\beta' & 1 & 0 \\ 0 & -1/3 & 1 \end{bmatrix} \begin{bmatrix} f_{t_m+1} \\ y^*_{t_m+1} \\ y^C_{t_m+1} \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Xi_{t_m+1} \end{bmatrix} \begin{bmatrix} f_{t_m} \\ y^*_{t_m} \\ y^C_{t_m} \end{bmatrix} + \begin{bmatrix} B\eta_{t_m+1} \\ 0 \\ 0 \end{bmatrix} \tag{9.48}$$

The aggregation rule (9.46) is implemented in a recursive way, by introducing a latent cumulator variable $y^C_{t_m} = \Xi_{t_m} y^C_{t_m-1} + \frac{1}{3} y^*_{t_m}$, where $\Xi_{t_m} = 0$ for t_m corresponding to the first month of the quarter and $\Xi_{t_m} = 1$ otherwise. The estimation of the model parameters follows Giannone et al. (2008).

Factor-MIDAS

It is possible to augment the MIDAS regressions with the factors extracted from a large dataset to obtain a richer family of models that exploit a large high-frequency dataset to predict a low-frequency variable.

While the basic MIDAS framework consists of a regression of a low-frequency variable on a set of high-frequency indicators, the Factor-MIDAS approach exploits estimated factors rather than single or small groups of economic indicators as regressors.

Marcellino and Schumacher (2010) propose alternative MIDAS regressions. In the standard MIDAS case, they follow Clements and Galvao (2008), while as a modification they evaluate a more general regression approach, labeled unrestricted Factor-MIDAS, where the dynamic relationship between the low-frequency variables and the high-frequency indicators is unrestricted, in contrast to the distributed lag functions as proposed by Ghysels et al. (2007). As a third alternative, they consider a regression scheme proposed by Altissimo et al. (2010), which considers only correlation at certain frequencies between variables sampled at high- and low- frequencies. This approach is called smoothed MIDAS, since the regression essentially eliminates high-frequency correlations.

The information set consists of a large set of stationary monthly indicators, X_{t_m}. The last observation is at time $T_m + w, w > 0$, allowing for at most $w > 0$ monthly values of the indicators that are earlier available than the lower-frequency variable to be estimated. X_{t_m} is modeled using a factor representation, where r factors F_{t_m} are estimated in order to summarize the information in X_{t_m}. The estimated factors, $\widehat{F}_{t_m}$, are used in the projection for the quarterly-frequency variable.

We now describe in details the three alternative Factor-MIDAS approaches proposed by Marcellino and Schumacher (2010), assuming again that the target variable is GDP. These approaches are tools for direct multi-step now- and forecasting, thus each model is for a specific forecast horizon.

The Basic Factor-MIDAS approach

In the basic Factor-MIDAS approach the explanatory variables used as regressors are estimated factors. Assume for simplicity $r = 1$, so that there is only one factor $\widehat{f}_{t_m}$. The Factor-MIDAS model for forecast horizon h_q quarters with $h_q = h_m/3$ is

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b\left(L_m;\theta\right)\widehat{f}^{(3)}_{t_m+w} + \varepsilon_{t_m+h_m}, \tag{9.49}$$

where $b\left(L_m;\theta\right) = \sum_{k=0}^{K} c\left(k;\theta\right) L_m^k$ and $c\left(k;\theta\right) = \frac{\exp\left(\theta_1 k+\theta_2 k^2\right)}{\sum_{k=0}^{K}\exp\left(\theta_1 k+\theta_2 k^2\right)}$.

$\widehat{f}^{(3)}_{t_m}$ is skip-sampled from the monthly factor $\widehat{f}_{t_m}$. Every third observation starting from the final one is included in the regressor $\widehat{f}^{(3)}_{t_m}$, i.e. $\widehat{f}^{(3)}_{t_m+w} = \widehat{f}_{t_m+w}, \forall t_m + w = ..., T_m + w - 6, T_m + w - 3, T_m + w$. Note that we take into account the fact that a monthly indicator is typically available also within the quarter for which no GDP figure is available. As described above in the MIDAS models, the exponential lag function provides a parsimonious way to consider monthly lags of the factors.

The model can be estimated using nonlinear least squares in a regression of y_{t_m} onto the factors $\widehat{f}^{(3)}_{t_m+w-h}$. The forecast is given by

$$y_{T_m+h_m|T_m+w} = \widehat{\beta}_0 + \widehat{\beta}_1 b\left(L_m;\widehat{\theta}\right)\widehat{f}^{(3)}_{T_m+w}. \tag{9.50}$$

The projection is based on the final values of estimated factors.

MIDAS regression can be generalized to more than one factor and extended with the addition of autoregressive dynamics. Details on factor estimation are provided in Section 9.3.2.

Smoothed MIDAS

A different way to formulate a mixed-frequency projection is proposed by Altissimo et al. (2010). The projection can be written as:

$$y_{T_m+h_m|T_m+w} = \widehat{\mu} + G\widehat{F}_{T_m+w} \tag{9.51}$$

$$G = \widetilde{\Sigma}_{yF}\left(h_m - w\right) \times \widehat{\Sigma}_F^{-1}, \tag{9.52}$$

where $\widehat{\mu}$ is the sample mean of GDP, G is the projection coefficient matrix, $\widehat{\Sigma}_F$ is the estimated sample covariance of the factors, and $\widetilde{\Sigma}_{yF}\left(k\right)$ is a particular cross-covariance with k monthly lags between GDP and the factors. $\widetilde{\Sigma}_{yF}\left(k\right)$ is not an estimate of the sample cross-covariance between factors and GDP, but a cross-covariance between smoothed GDP and factors. The smoothing aspect is introduced in $\widetilde{\Sigma}_{yF}\left(k\right)$ as follows. The estimated covariance between $\widehat{F}_{t_m-k}$ and y_{t_m} is:

$$\widehat{\Sigma}_{yF}\left(k\right) = \frac{1}{T^*-1}\sum_{t_m=M+1}^{T_m} y_{t_m}\widehat{F}^{(3)\prime}_{t_m-k}, \tag{9.53}$$

where T^* is the number of observations available to compute the cross-covariances. Note the use of skip-sampled factors, since GDP is available only quarterly. Given $\widehat{\Sigma}_{yF}\left(k\right)$, the estimated cross-spectral matrix is

$$\widehat{S}_{yF}\left(\omega_j\right) = \sum_{k=-M}^{M}\left(1 - \frac{|k|}{M+1}\right)\widehat{\Sigma}_{yF}\left(k\right)e^{-i\omega_j k}, \tag{9.54}$$

at frequencies $\omega_j = \frac{2\pi j}{2H}$ for $i = -H, ..., H$ using a Bartlett lead-lag window. The low-frequency relationship between $\widehat{F}_{t_m-k}$ and y_{t_m} is obtained by filtering out cross fluctuations at frequency higher than a certain threshold π/q, using the frequency-response function $\alpha\left(\omega_j\right)$,defined as $\alpha\left(\omega_j\right) = 1, \forall\,|\omega_j| < \pi/q$ and zero otherwise. The autocovariance matrix $\widetilde{\Sigma}_{yF}\left(k\right)$ reflecting low-frequency co-movements between $\widehat{F}_{t_m-k}$ and

y_{t_m} is obtained by inverse Fourier transform:

$$\widetilde{\Sigma}_{yF}(k) = \frac{1}{2H+1} \sum_{j=-H}^{H} \alpha(\omega_j) \widehat{S}_{yF}(\omega_j) e^{i\omega_j k}. \tag{9.55}$$

Note that $\widehat{\Sigma}_{yF}(k)$ is a consistent estimator of the true cross-covariance, if the sample size is sufficiently large.

Unrestricted MIDAS

An alternative to the two previous models is the unrestricted lag order specification, as described in the above Section 9.2.3:

$$y_{T_m+h_m} = \beta_0 + D(L_m) \widehat{F}^{(3)}_{t_m+w} + \varepsilon_{t_m+h_m}, \tag{9.56}$$

where $D(L_m) = \sum_{k=0}^{K} D_k L_m^k$ is an unrestricted lag polynomial of order K.

$D(L_m)$ and β_0 are estimated by OLS. To specify the lag order in the empirical application, Marcellino and Schumacher consider a fixed scheme with $k = 0$ and an automatic lag length selection using the BIC.

9.3 Ragged-edge data

After having analyzed the various techniques to deal with mixed-frequency data, in this section we review the estimation methods that can handle ragged-edge data, that is datasets which are not balanced, due to the presence of missing values at the end of the sample for some indicators.

First, we discuss the issues of estimation and forecasting MF-VAR in the presence of missing observations due to publication lags and to the low-frequency nature of one variable. We follow Mariano and Murasawa (2010) in the discussion of estimation of the state-space form and forecasting with the use of Kalman filter and/or smoother.

Going further, we analyze issues related to factor estimation in the presence of ragged-edge data. Marcellino and Schumacher (2010) review three different methods to tackle it. First, the method proposed by Altissimo et al. (2010), who realign each time series in the sample in order to obtain a balanced dataset, and then estimate the factors with dynamic PCA. As an alternative, to consider missing values in the data for estimating factors Stock and Watson (2002) propose an EM algorithm together with the standard principal component analysis (DPCA). As a third method, Doz et al. (2011) propose a factor estimation approach based on a complete parametric representation of the large factor model in state-space form.

9.3.1 Estimating the MF-VAR with missing observations

As already anticipated in Section 9.2, Kalman filtering techniques can handle ragged-edge data and missing values due to publication lags and the low-frequency nature of a time series.

Estimation

The state-space representation of the mixed-frequency VAR model is described by equations (9.26) and (9.27), reported also here:

$$\begin{aligned} s_{t_m} &= F s_{t_m-1} + G v_{t_m} \\ \begin{pmatrix} y_{t_m} - \mu_y \\ x_{t_m} - \mu_x \end{pmatrix} &= H s_{t_m}. \end{aligned}$$

It can be estimated by maximum-likelihood even in the presence of missing observations due to publication lags and the low-frequency nature of GDP. However, as Mariano and Murasawa (2010) mention in their paper, when the number of parameters is large, the ML method can fail to converge.

In these cases, it is useful to implement the EM algorithm modified to allow for missing observations. Mariano and Murasawa (2010) consider the missing values as realizations of some iid standard normal random variables, i.e.

$$y_{t_m}^{+} = \begin{cases} y_{t_m} & \text{if } y_{t_m} \text{ is observable} \\ \zeta_{t_m} & \text{otherwise} \end{cases} \tag{9.57}$$

where ζ_{t_m} is a draw from a standard normal distribution independent of the model parameters.

The measurement equation is modified accordingly in the first two months of each quarter, where the upper row of H is set to zero and a standard normal error term is added, so that the Kalman filter skips the random numbers. Since the realizations of the random numbers do not matter in practice, the authors suggest to replace the missing values with zeros. Then, the EM algorithm is employed to obtain estimates of the parameters.

Estimation of latent monthly real GDP

Mariano and Murasawa (2010) use the Kalman smoother instead of the Kalman filter, because it uses more information and also simplifies the formulation of the state-space model. Although GDP growth for a particular month is not available, the smoother considers the monthly indicators available for the same quarter, so that nowcasting is also possible. For the months in which no observations are available also for the monthly indicators, the Kalman smoother acts exactly as the Kalman filter.

9.3.2 Estimating the factors with ragged-edge data

Factor forecasting with large, single-frequency data is often carried out using a two-step procedure. First, the factors are estimated and, second, a dynamic model for the variable to be predicted is augmented with the estimated factors. The same procedure can be used also in case of mixed-frequency data. As an alternative, factor estimation and forecasting can be conducted in a single step, in the contest of a parametric model.

The literature provides various ways to estimate the factors, in case of balanced datasets. However, in the following subsections we describe the methods that can handle ragged-edge data, that is datasets which are not balanced due to the presence of missing values at the end of the sample for some indicators.

Vertical realignment of data and DPCA

Altissimo et al. (2010) provide a convenient way to solve the ragged-edge problem. They propose to realign each time series in the sample in order to obtain a balanced dataset. Assume that variable i is realized with k_i months of publication lags. Thus, given a dataset in period $T_m + w$, the final observation of time series i is for period $T_m + w - k_i$. Altissimo et al. (2010) propose to realign the series in this way:

$$\widetilde{x}_{i,T_m+w} = x_{i,T_m+w-k_i}, \tag{9.58}$$

for $t_m = k_1 + 1, ..., T_m + w$. Once applied to each time series, the result is a balanced dataset $\widetilde{X}_{t_m}$, for $t_m = \max\{k_1\}_{i=1}^{N} + 1, ..., T_m + w$.

Given this balanced dataset, Altissimo et al. (2010) propose dynamic PCA to estimate the factors. The two-step estimation procedure introduced by Forni et al. (2005) directly applies, since the dataset is balanced.

One of the main advantages of this method is the simplicity. A drawback is that the availability of data determines cross-correlations between variables. Moreover, data releases are not the same over time, so that dynamic correlations within the data change and, as a consequence, factors change over time. The same happens if factors are reestimated at a higher frequency than the one of the factor model, for example in the case of a monthly factor model reestimated several times within a month, in correspondence of new releases of the data.

EM algorithm and PCA

As an alternative, to consider missing values in the data for estimating factors, Stock and Watson (2002) propose an EM algorithm combined with the standard PCA. Call X_i the column i of the dataset X_{t_m}: not all observations are available, due to publication lags. The vector X_i^{obs} contains the observations available for variable i, a subset of X_i. More precisely, the relation between observed and not fully observed variables is

$$X_i^{obs} = A_i X_i, \tag{9.59}$$

where A_i is the matrix that tackles missing values. A_i is equal to the identity matrix, in case there are no missing values in the series. When an observation is missing at the end of the sample, the corresponding final row of the identity matrix is removed.

The EM algorithm consists in the following steps:

1. Provide an initial guess $\widehat{X}_i^{(0)}, \forall i$. These guesses together with the fully available series provide a balanced dataset $\widehat{X}^{(0)}$. With a balanced dataset, standard PCA gives initial monthly factors $\widehat{F}^{(0)}$ and loadings $\widehat{\Lambda}^{(0)}$.
2. E-step: an updated estimate of the missing observations for variable i is provided by the expectation of X_i conditional on X_i^{obs}, factors $\widehat{F}^{(j-1)}$ and loadings $\widehat{\Lambda}^{(j-1)}$ from the previous iteration

$$\widehat{X}_i^{(j)} = \widehat{F}^{(j-1)}\widehat{\Lambda}_i^{(j-1)} + A_i' \left(A_i' A_i\right)^{-1} \left(X_i^{obs} - A_i \widehat{F}^{(j-1)}\widehat{\Lambda}_i^{(j-1)}\right). \tag{9.60}$$

   We can recognize two components in the update: the common component from the previous iteration $\widehat{F}^{(j-1)}\widehat{\Lambda}_i^{(j-1)}$, and a low-frequency idiosyncratic component $X_i^{obs} - A_i \widehat{F}^{(j-1)}\widehat{\Lambda}_i^{(j-1)}$, distributed by the projection coefficient $A_i' \left(A_i' A_i\right)^{-1}$ on the high-frequency periods, see Breitung and Schumacher (2008).
3. M-step: repeat the E-step for each variable i, in order to obtain a balanced dataset. Reestimate all the factors $\widehat{F}^{(j)}$ and loadings $\widehat{\Lambda}^{(j)}$ by PCA. Go back to step 2 until convergence.

After convergence, the EM algorithm provides both the monthly factor estimates $\widehat{F}_{t_m}$ and the estimates of the missing values of the time series, see also Angelini et al. (2006).

Large parametric factor model in state-space form

Doz et al. (2011) propose a factor estimation approach based on a complete representation of the large-factor model in state-space form. The full state-space model has the form

$$X_{t_m} = \Lambda F_{t_m} + \xi_{t_m} \tag{9.61}$$

$$\Psi\left(L_m\right) F_{t_m} = B\eta_{t_m}. \tag{9.62}$$

Equation (9.61) is the static factor representation of X_{t_m}. Equation (9.62) specifies a VAR structure of the factors, with lag polynomial $\Psi\left(L_m\right) = \sum_{i=1}^{p} \Psi_i L_m^i$. $\eta_{t_m}$is a $q-$dimensional vector that contains the orthogonal dynamic shocks that drive the r factors. The factors F_{t_m} represent the states, while ξ_{t_m} is the stationary idiosyncratic component which admits a Wold representation. The shocks driving the factors and the idiosyncratic components are assumed to be independent. If the X_{t_m} is of a small dimension, the model can be estimated by iterative maximum likelihood. If the dimension is large, iterative ML is infeasible, so Doz et al. (2011) propose a quasi-ML to estimate the factors. For a given number of factors, r, and dynamic shocks, q, the estimation follows the steps illustrated below:

1. Estimate $\widehat{F}_{t_m}$ using PCA as an initial estimate. The estimation is based on the balanced part of the data, obtained by removing the values at the end of the sample that create the unbalancedness.

2. Estimate the loadings $\widehat{\Lambda}$ by regressing X_{t_m} on the factors estimated in the previous step, $\widehat{F}_{t_m}$. Estimate also the covariance of the idiosyncratic components $\widehat{\xi}_{t_m}$, denoted as $\widehat{\Sigma}_{\xi}$.

3. Estimate a VAR(p) on the factors $\widehat{F}_{t_m}$, obtaining $\widehat{\Psi}(L_m)$, and $\widehat{\Sigma}_{\varsigma}$, the residual covariance of $\widehat{\varsigma}_{t_m} = \widehat{\Psi}(L_m)\widehat{F}_{t_m}$.

4. To obtain an estimate of B, given the number of dynamic shocks q, apply an eigenvalue decomposition of $\widehat{\Sigma}_{\varsigma}$.Call M the $(r \times q)$ −dimensional matrix of the eigenvectors corresponding to the q largest eigenvalues, and call P the $(q \times q)$ −dimensional matrix with the largest eigenvalues on the main diagonal and zero otherwise. The estimate of B is $\widehat{B} = M \times P^{-1/2}$. All the parameters and coefficients in the system of equations (9.61) and (9.62) are then fully specified. The model is cast into state-space form.

5. The Kalman filter or smoother yields new estimates of the monthly factors. The dataset used now is the unbalanced one, where T_m is the last observation available in the whole set of monthly series. The Kalman filter also provides estimates and forecasts for the missing values conditional on the model structure and properties of the shocks.

Note that the coefficients in the system have to be estimated from a balanced sub-sample of data, as in step 1 there is the need of a fully balanced dataset for PCA initialization. Nevertheless, in step 5 the factor estimation based on the Kalman filter applies to the unbalanced dataset. The solution is to estimate the coefficients outside the state-space model and avoid to estimate a large number of coefficients by iterative ML.

9.4 A comparison of the different methods

So far, we have seen that several methods have been proposed in the literature to deal with mixed-frequency data, possibly with a ragged edge structure. In general, there is an agreement on the fact that exploiting data at different frequencies matters for nowcasting and short term forecasting. We now try to summarize the advantages and disadvantages of the different methods, comparing their most important features.

Bridge equations are still one of the most used techniques, especially in short run forecasting, because they are pretty easy to estimate and interpret, and allow computing early estimates of the low-frequency variable. The drawback is that they are purely statistical models, where the regressors are included only because they contain timely updated information. Therefore, if the model that exploits the high-frequency information is misspecified, the error transmits to the bridge equation and to the forecasts that are obtained recursively.

A more sophisticated way to deal with data sampled at different frequencies is the state-space approach. Casting the model in state-space form has the advantage of jointly specifying the dynamics of the indicators and of the variable to be explained without imposing any a-priori restriction. Moreover, since the low-frequency series is seen as a high-frequency series with missing values, the use of the Kalman filter permits the estimation of these missing data. As shown in Bai et al. (2011), the Kalman filter results to be the optimal filter in population, when ignoring parameter estimation errors and assuming that the model is correctly specified. Therefore, under these ideal circumstances, the state-space approach cannot be beaten by any other method. On the other side, there are also some drawbacks from the use of this approach: first of all, in most of the cases it is computationally complex, and the complexity increases with the number of variables involved, so that most of the time only small-scale models can be estimated. Moreover, the state space approach requires the correct specification of the model in high frequency, which is even more complex than usual given the missing observations in the dependent variable.

An alternative way to deal with mixed-frequency data is the MIDAS approach. Even though in population, when the process is correctly specified, MIDAS is coarse than the optimal Kalman filter, it can be more robust in the presence of mis-specification. Moreover, the lag polynomials are based on a very small number of parameters, allowing the MIDAS models to be parsimonious, even though it is still not clear which is the best polynomial specification. Contrary to what stated for the state-space models, MIDAS models can be easily

estimated by NLS. However, it is only possible to obtain a high frequency update of the expected low frequency realization, not an estimate of the missing values in the low frequency variable.

Both the state-space and the MIDAS approaches can be combined with a factor specification, in order to use the information in a large dataset, possibly with a ragged edge. Whether factor methods provide more precise estimates and forecasts than VARs or single equation methods is a matter for empirical investigation, since there is a trade-off between model complexity and extended information set.

9.5 An overview of empirical studies

In this section we review the empirical literature on forecasting with mixed-frequency and ragged-edge data, providing some examples of all the models and estimation methods outlined in the previous sections.

9.5.1 Bridge equations

Bridge equations have been one of the first methods employed in nowcasting the current state of the economy, making use of the monthly information available. Studies of this kind have been conducted for the nowcasts of different economies. A common finding of these studies is that the exploitation of intra-period information reduces the forecasting error in the majority of the cases. The applications concern both “supply-side” and “demand-side” models.

Looking at US data, Ingenito and Trehan (1996) construct a model that predicts current quarter real GDP based on knowledge of nonfarm payrolls, industrial production and real retail sales, which have the advantage of being released at a monthly frequency by the middle of the following month. In order to produce a model that predicts real GDP, the authors rely on auxiliary models that generate forecasts of the indicator variables themselves. Their evidence shows that consumption data provide key information about current output, and that retail sales release allows to have a good forecast of contemporaneous consumption.

Stark (2000) presents evidence on the usefulness of conditioning quarterly model forecasts on monthly current-quarter data, in the case of the US economy. Starting by generating a one-step-ahead forecast from a quarterly Bayesian vector error correction model, the author then specifies a monthly statistical model for variables that are thought to carry information about each of the variables in the quarterly model and uses it to generate sequences of current-quarter quarterly-average forecasts from the monthly indicators. Once he has these quarterly-average monthly indicator forecasts, he forms updated estimates of the quarterly model’s current-quarter forecast. The findings show that exploiting monthly information produces economically and statistically significant improvements, particularly large especially during periods of recession.

A study by Barhoumi et al. (2011) presents a model to predict French gross domestic product (GDP) quarterly growth rate. The authors employ the bridge equations to forecast each component of the GDP, and select the monthly explanatory variables among a large set of hard and soft data. They find that changing the set of equations over the quarter is superior to keeping the same equations over time. These models turn out to beat the benchmark in terms of forecasting performance.

Studies involving bridge equations can be found for many other countries. In particular, bridge models have been employed also for nowcasting Euro Area GDP growth. As an example, we consider Baffigi et al. (2004). In this paper, bridge models are estimated for aggregate GDP and components, both area-wide and for the main countries of the Euro Area. Their short-term performance is then assessed with respect to benchmark univariate and multivariate standard models, and a small structural model. The results shown in the paper are clear-cut: bridge models performance is always better than benchmark models, provided that at least some indicators are available over the forecasting horizon. As far as the type of aggregation is concerned, the supply-side approach (modelling aggregate GDP) performs better than the demand-side approach (aggregation of forecasts by national account components). The supply-side models highlight the

role of industrial production and manufacturing surveys as the best monthly indicators. Looking at the demand-side models, from the different equations estimated in this paper, private consumption results well tracked by retail sales index, while the consumer confidence index plays a minor role; in the case of investment a major role seems to be played by survey variables.

Diron (2008) makes use of bridge equations with Euro Area data to provide an assessment of forecast errors, which takes into account data-revisions. Using four years of data vintages, the paper provides estimates of forecast errors for Euro Area real GDP growth in genuine real-time conditions and assesses the impact of data revisions on short-term forecasts of real GDP growth. Given the size of revision to indicators of activity, the assessment of reliability of short-term forecasts based on revised series could potentially give a misleading picture. Nevertheless, averaging across all bridge equations, forecasts of individual quarters tend to be similar whether they are based on preliminary or revised data. More specifically, the RMSEs based on real-time and pseudo real-time exercises are quite similar and both smaller compared with AR forecasts of GDP, considered as the benchmark. The difference in forecast accuracy is significant according to Diebold and Mariano tests, highlighting that short-term forecasts based on bridge equations are informative. Moreover, the paper investigates the contributions of the various sources to the overall forecasting errors. Revisions to the monthly variables and to GDP growth account only for a small share of the overall forecast error, while the main sources are from extrapolation of the monthly indicators. The relative contributions of extrapolation and of revision of monthly indicators vary depending on whether the equations include hard data, in which case both sources are significant, or survey and financial variables, in which case these two sources tend to have a smaller weight.

In the context of nowcasting, it has become more common to exploit the information coming from a large set of variables. Therefore, recent studies combine the bridge models with factors, in what in is called “bridging with factors”. This new kind of model is related to the one described in Section 9.2.5 by Giannone et al. (2008). In Section 9.5.4, we will review these studies and compare the performance of this new kind of bridge with factors to the standard bridge models and other benchmarks.

9.5.2 MIDAS models

In the first applications, MIDAS models have been applied to financial data, investigating the relation between the conditional mean and the conditional variance of the stock market returns or future volatility, see Ghysels et al. (2005) as an example. Clements and Galvao (2008) are the first to apply MIDAS regressions to macroeconomic data. In the next paragraphs we will overview applications based on financial and/or macroeconomic data.

Ghysels et al. (2005)) study the intertemporal relation between the conditional mean and the conditional variance of the aggregate market return. In support of Merton’s ICAPM, the authors find a positive significant and robust relation between risk and return. They also find that the MIDAS estimator is a better forecaster of the stock market variance than two other benchmark models: rolling window and GARCH estimators. The authors also focus on the asymmetric reaction of volatility to positive and negative shocks. They find that positive shocks have a bigger impact overall on the conditional mean of returns, are slower to be incorporated in the conditional variance, and are much more persistent, while negative shocks have a large initial, but temporary, effect on the variance of returns. Ghysels et al. (2006) consider various MIDAS regressions to predict volatility in a parsimonious way with data at different frequency. They find that daily realized power is the best predictor of future increments in quadratic variation. Surprisingly, the direct use of high-frequency (5 minutes) data does not improve volatility prediction.

Ghysels et al. (2009) compare three different approaches of producing multi-period-ahead forecasts of volatility: iterated, direct and MIDAS. The comparison is conducted out-of-sample using returns data of the US stock market portfolio and a cross section of size, book-to-market and industry portfolios, in terms of the average forecasting accuracy, using the MSFE. The direct approach provides the worst forecasts. Iterated forecasts are suitable for shorter horizons, while MIDAS forecasts perform well at long horizons.

Clements and Galvao (2008) introduce the use of MIDAS regressions in forecasting macroeconomic data. They also look at whether a mixed-data sampling approach including an autoregressive term can improve forecasts of US real output growth. They conduct a real-time forecasting exercise that exploits monthly vintages of the indicators and the quarterly vintages of the output growth, consistent with the timing of the releases of the different data vintages. The authors find that the use of within-quarter information on monthly indicators can result in a marked reduction in RMSE compared with the more traditional quarterly-frequency AR or AR distributed-lag models. Moreover, Clements and Galvao (2009) evaluate the predictive power of leading indicators for output growth up to one year, using MIDAS approach to combine multiple leading indicators in a parsimonious way. The results confirm that MIDAS is a useful instrument to improve forecasts. Moreover, they show that the use of real-time vintage data improves forecast performance and that the predictive power of the indicators is stronger when the aim is to forecast final data rather than first-released data, although first releases can generally be forecasted more accurately.

Foroni et al. (2014) compare the performance of the MIDAS with functional distributed lags estimated with NLS to the one of the U-MIDAS, the unrestricted version analyzed in Section 9.2.3. In Monte Carlo experiments, they show that U-MIDAS generally performs better than MIDAS when mixing quarterly and monthly data. On the other hand, with larger differences in sampling frequencies, distributed lag-functions outperform unrestricted polynomials. In an empirical application on out-of-sample nowcasting GDP in the Euro area and the US using monthly predictors, they find a good performance of U-MIDAS for a number of indicators, albeit the results depend on the evaluation sample.

In the recently increasing literature, which is exploiting the availability of a huge number of financial time series on a daily basis to forecast macroeconomic time series, the empirical evidence in support of the use of high-frequency financial series is rather mixed. On the one side, it is useful to use this great amount of timely information, but on the other side there is a question on how to weight the daily observations and to filter these data, to get rid of the possible noise. Results from recent studies suggest that daily variables seem to have useful information for forecasting inflation and economic activity.

Among these studies, Ghysels and Wright (2009) propose methods for using asset price data to construct daily forecasts of upcoming survey releases, employing MIDAS regression models and a more structural approach based on the Kalman filter to estimate what forecasters would predict if they were asked to make a forecast each day, treating their forecasts as missing data to be interpolated. Their aim is to obtain high-frequency measures of forecasters' expectations. The authors consider two surveys in their empirical work: the Survey of Professional Forecasters and the Consensus Forecasts, and use the daily asset prices to predict the upcoming releases of either of the two surveys. In an in- and out-of-sample forecasting exercise, both approaches (MIDAS and Kalman filter) perform better than the simple random walk benchmark forecasts.

Andreou et al. (2010) assess whether daily financial data can improve macroeconomic forecasting, employing MIDAS regression models. They forecast US quarterly inflation rate and economic growth using a dataset including daily, monthly and quarterly indicators. An important advantage of the MIDAS model is that it can provide new forecasts as daily data become available. The authors find that on average daily financial predictors improve the forecasts of quarterly inflation and GDP relative to the AR benchmark model.

Monteforte and Moretti (2010) present a mixed frequency model for daily forecasts of Euro area inflation in real-time. The model they use allows to combine a monthly core inflation estimated from a dynamic factor model with daily financial market variables, which provide timely information on the recent shocks. They compare the results of this mixed-frequency model with standard univariate and multivariate models with monthly data, and also with the forecasts implied by financial market expectations extracted from future contracts. In both cases, the mixed frequency approach shows a superior predictive power.

9.5.3 Mixed-frequency VAR models

As we have seen in Section 9.2.4, studies on MF-VAR models have been conducted both in a classical and in a Bayesian context. We now outline the main empirical studies conducted in both frameworks.

Mittnik and Zadrozny (2005) evaluate a Kalman-filtering-based maximum- likelihood estimation method for forecasting German real GDP at monthly intervals. They estimate a VAR(2) model of quarterly GDP and up to three monthly indicator variables (industrial production, current and expected business conditions). They find that in general monthly models produce better short-term GDP forecasts, while quarterly models produce better long-term GDP forecasts.

Mariano and Murasawa (2010) apply the MF-VAR method to construct a new coincident indicator, i.e. an estimate of monthly real GDP. What they find is that the coincident index based on the VAR model is close to the one obtain by a factor model, and they both track well quarterly real GDP, although they are quite volatile.

Kuzin et al. (2011) compare the MF-VAR, as presented in Mariano and Murasawa (2010), with the MIDAS approach to model specification in the presence of monthly and quarterly series. MIDAS leads to parsimonious models, while MF-VAR does not restrict the dynamics but suffers from the curse of dimensionality. The authors argue that it is difficult to rank the different approaches a priori, so they compare their performance empirically, considering an AR process as a benchmark. The two approaches tend to be more complementary than substitutes, since the MF-VAR performs better for longer horizons, whereas MIDAS for shorter horizons. Looking at the relative MSE of the different models with respect to the benchmark, the mixed-frequency models perform relatively well, especially when forecast combinations are adopted.

Similar evidence is found by Foroni and Marcellino (2014a) in their paper which focuses on different methods proposed in the literature to deal with mixed-frequency and ragged-edge datasets. The authors discuss the performance of the different methods on now- and forecasting the quarterly growth rate of the Euro Area GDP and its components, using a very large set of monthly indicators. They also find that MF-VAR outperforms the MIDAS approach only at longer horizons.

Ghysels (2011) introduces a different MF-VAR representation, in which he constructs the MF-VAR process as stacked skip-sampled processes. In this paper, the author characterizes explicitly the mis-specification of a traditional low frequency VAR and the consequent misspecification in the impulse response functions. Moreover, since the MF-VAR specified in this way can also characterize the timing of information releases, he shows how Choleski factorizations are a more natural tool for impulse response analysis because the elements in the vector represent a sequence of time events. As another contribution, he studies a Bayesian approach which accommodates the potentially large set of parameters to be estimated.

One of the earliest studies on Bayesian estimation of MF-VAR is the paper by Chiu et al. (2011). In this paper, the authors develop a Gibbs sampling approach to estimate a VAR with mixed and irregularly sampled data. The focus of the paper is on the parameter estimation. In an exercise with simulated data, the authors show that taking into account mixed-frequency data allows to obtain smaller root mean squared errors for all the parameter estimates regardless of the sample size and of the correlation between the variables of the system. These results find confirmation also in the two empirical examples, conducted with data respectively at monthly and quarterly and weekly and monthly frequency, for which the authors compare the posterior distributions of the parameters and the impulse response functions.

Another study by Viefers (2011) reconsiders the estimation of a MF-VAR as in Mariano and Murasawa (2010). First, the author makes use of the Bayesian MCMC algorithm to simulate and estimate the model, and second he extends the MF-VAR to allow for regime switching. In his model, the inference is based on the joint posterior density of all the unknowns. The findings of the simulation study suggest that inference on the latent series and the regime processes is fairly precise, although there is a more pronounced imprecision in the estimation of the VAR parameters. In the empirical exercise, the author considers monthly and quarterly data for the US economy. The results on the regime probabilities show a relative high ability to identify the same recession dates provided by the NBER, although the probabilities tend to be more erratic and much worse in the most recent years.

Schorfheide and Song (2011) conduct a forecasting exercise on US data exploiting MF-VAR models. The goal of their paper is to study the extent to which the incorporation of monthly information improves the forecasts compared to models based on quarterly aggregated data. The analysis is conducted for 11 US variables, of

which 3 observed at quarterly level, in a real-time context. The authors find that the monthly series provide important information in the short run, with significant RMSE reductions obtained with the mixed-frequency model. Moreover, the more intra-quarter information is available, the increasing the improvements.

9.5.4 Factor models

Applications of small-scale factor models

Small-scale factor models have been frequently employed in the literature to construct a coincident indicator, which is able to track the development of the economy in real-time. In what follows, we describe the main studies which employ small-scale factor models which extract an index and provide, in some of the cases presented, short-term forecasts of the real GDP growth.

As described also in section 9.2.5, Mariano and Murasawa (2003) propose a new coincident index of business cycles that relies on both monthly and quarterly indicators. Stock and Watson (1989) construct a coincident index by applying maximum likelihood (ML) factor analysis to the four monthly coincident indicators. Mariano and Murasawa extend the Stock–Watson coincident index by including quarterly real GDP and compare the turning point detection performance of the two indices. What they find is that the behavior of the common component is quite different from monthly real GDP, and more generally that the behavior of the common factor depends on the choice of the component indicators and therefore the monthly real GDP and the common factor component can have different turning points.

A different application of the Mariano and Murasawa model can be found in Frale et al. (2011). This paper proposes a new monthly indicator of the euro area economic conditions, EUROMIND, based on tracking real GDP on a monthly basis. The construction of this new monthly indicator of GDP is carried out indirectly through the temporal disaggregation of the value added by supply sectors from the output side and at the same time through the temporal disaggregation of the main components of the demand from the expenditure side. The two estimates are combined with optimal weights reflecting their precision. Therefore, the indicator is based on information at both monthly and quarterly level, modelled with a dynamic factor specification cast in state-space form, where computational efficiency is achieved by implementing univariate filtering and smoothing procedures, which also allows to handle the ragged-edge problem and other data irregularities in a unified framework. The authors find satisfactory results in the application of the model to the sectorial data, while the results are less convincing on the expenditure side. In a second paper, Frale et al. (2010) introduce a modification in the model which consists of the introduction of a second common factor, capturing the contribution of the survey variables as coincident indicators. What they find is that the second factor loads significantly on the survey variables for the industry sector and for exports. Moreover, they also attempt to isolate the news content of each block of series by using a real-time database: the analysis of the revisions in the data indicates that the contribution of surveys is not negligible.

Camacho and Perez-Quiros (2010) construct a different coincident indicator of the Euro area economy, the so-called Euro-STING indicator, which evolves accordingly to the Euro area dynamics and it is also based on an extension of the dynamic factor model described in Mariano and Murasawa (2003). The authors accommodate the GDP releases (flash, first and second estimate) in a statistical model to examine the impact of preliminary announcements and data revisions in the accuracy of real time forecasting. They assume that monthly growth rates of quarterly series and monthly growth rates of hard indicators have a direct relation with the common factor, which represents the common component that drives the series dynamics. However, they model the relation of the common factor with the soft indicators in a different way, and precisely they relate the level of the soft indicators considered with the year-on-year common growth rate, written as the sum of current values of the common factors and the last eleven lagged values. In their empirical application, they deal with a relatively small number of indicators. What they find is that exploiting information within each quarter their model improves upon the accuracy of preliminary announcements in forecasting GDP and forecasting uncertainty decreases during the forecasting period. Moreover, not only hard indicators are useful in forecasting GDP but also business surveys are relevant, especially in the months when real activity data are not available yes due to publication lags.

What Camacho and Perez-Quiros (2010) do for the Euro area is closely related to the empirical work done by Evans (2005) for the US, who applies a model that allows for variable reporting lags and temporal aggregation to a wide range of US macroeconomic data releases. The author models the growth in GDP as the quarterly aggregate of an unobserved daily process and specifies the relationship between GDP, data releases on GDP growth and on other macroeconomic variables in such a way to accommodate the different timing of data releases. By writing the model in state-space form (similar to Mariano and Murasawa (2003) but accommodating for a more complex timing of the releases), Evans (2005) obtains a real-time estimate of GDP on a daily basis as a product of the Kalman filter applied to estimate the model. The results seem promising, showing that within quarter data releases contain useful information for real-time estimation and forecasting of GDP. However, gaps between the real-time estimates and ex-post GDP data remain both persistent and significant.

Another extension of the small-scale factor model of Mariano and Murasawa (2003) has been analyzed by Marcellino et al. (2013), to investigate business cycle dynamics and for forecasting GDP growth at short-term horizons in the euro area. While so far the parameters of the model have been considered as constant, the authors consider time variation in the variance of the shocks, and they generalize the setup of Mariano and Murasawa (2003) to allow for continuous shifts in the volatility of the shocks both to the common components and to the single indicators. To do so, they model volatility shifts as independent random walks. Moreover, differently from the other studies, the model is estimated with a Bayesian technique, using a Gibbs sampling procedure. The authors use the model to evaluate the impact of macroeconomic releases on point and density forecast accuracy and on the width of forecast intervals, and they show how their setup allows to make a probabilistic assessment of the contribution of releases to forecast revisions. From a pseudo out of sample forecasting exercise, they find that stochastic volatility contributes to an improvement in density forecast accuracy.[4]

Bridging with factors

As we have seen in Section 9.2.5, Giannone et al. (2008) provide a framework that formalizes the updating of the nowcast and forecast of output and inflation as data are released throughout the month and that can be used to evaluate the marginal impact of new data releases on the precision of the now- and forecasts as well as the marginal contribution of different groups of variables. The framework they propose is adapted from the parametric dynamic factor model in state-space form proposed by Doz et al. (2011) that helps handling ragged-edge data. They extract monthly factors and use them in a state-space framework to forecast monthly GDP. The authors construct pseudo intra-month vintages according to a stylized data release calendar. As a new block of information is released, the factors are reestimated and the nowcast updated. The main finding is that information matters: the precision of the signal increases monotonically within the month, with the release of new data. Timeliness of the release and quality matters for decreasing uncertainty. Banbura et al. (2012) present a similar but updated and extended application that confirms this finding.

Barhoumi et al. (2008) compare small bridge equations and forecast equations in which the bridging between monthly and quarterly data is achieved through a regression on factors extracted from large monthly datasets. The authors consider the framework proposed by Giannone et al. (2008), but they also extract the factors following Forni et al. (2005), using generalized principal components which allow to take into account the ragged-edge structure of the dataset. In their paper, they focus on the Euro Area as a whole as well as on single Euro Area countries. The results obtained for the Euro Area countries show that models that exploit timely monthly releases fare better than quarterly models, and among those, factor models do generally better than averages of bridge equations.

[4] Aastveit et al. (2011) develop a nowcasting system that combines forecasts from VAR models, bridging equations, and factor models. They also focus on density forecasts, for which their approach works well.

Factor models in a mixed-frequency state space representation

Enriching the model proposed by Giannone et al. (2008), Banbura and Rünstler (2011) and Banbura and Modugno (2010) make use of a large state-space model allows for joint estimation of GDP and the factors in a single framework.

Banbura and Rünstler (2011) develop measures for understanding the importance of an individual series in the forecasts: they derive the weights of the series in the forecast and use them to calculate their contributions to the forecasts. Moreover, they assess the gains in forecast precision due to certain series by measuring the increase in uncertainty once the series have been removed from the explanatory variables. Banbura and Rünstler (2011) use a factor model which implements the common factors as unobserved components in a state-space form, and integrate the monthly factor model and a forecast equation for quarterly GDP in a single state-space representation, using a mixed-frequency setup. The authors confirm the finding of the importance of intra-quarter information, showing that a quarterly AR model is clearly outperformed by the factor model. Moreover they find evidence that differences in the timeliness of data releases can have strong effects on the optimal weights of individual series in the forecast and on their contribution to forecast precision.

Banbura and Modugno (2010) extend the analysis in Banbura and Rünstler (2011) by augmenting the dataset by short history indicators and quarterly series. Moreover, they allow the model to have an AR(1) idiosyncratic component. In their pseudo real-time exercise, they recursively forecast the Euro area GDP on the basis of a large monthly dataset. Compared to the previous data employed in Banbura and Rünstler (2011), they introduce short-history indicators, the Purchasing Managers' Surveys, available only from mid 1997. The results obtained including these new short history monthly indicators are similar to the results obtained without including them, therefore it seems that these additional series do not improve the precision of the projections. Also with the explicit modelling of the serial autocorrelation of the idiosyncratic component the results do not improve significantly.

Moreover, Angelini et al. (2008) provide an out-of-sample evaluation of the method presented by Banbura and Rünstler (2011), and compare the forecasting performance of this approach to the one obtained by pooling the forecasts from different selected bridge equations. In order to evaluate the impact of new data releases on current GDP nowcast throughout the quarter, they update the model two times per month, measuring the accuracy of the forecasts computed using the information available at each date. The results they find show that the factor model forecast tracks GDP more accurately, most likely due to the fact that the factors take into account the information content of cross correlations across series.

Factor-MIDAS

Marcellino and Schumacher (2010) propose to merge factor models with the MIDAS approach, which allows to now- and forecast low frequency variables as GDP exploiting information in a large set of higher-frequency indicators. They consider three different MIDAS approaches - basic, smoothed and unrestricted - and the three alternative factor estimation methods that can account for unbalanced datasets, explained in Section 9.3.2, to have a total of nine Factor-MIDAS approaches. They then focus on German GDP and conduct now- and forecasting of quarterly GDP growth with a large set of timely monthly economic indicators. To relate Factor-MIDAS to the methods from the existing literature, the authors introduce two other approaches in the empirical comparison: a single-frequency factor model based on quarterly aggregated data and the integrated state-space approach by Banbura and Rünstler (2011). In terms of empirical results, MIDAS with exponentially distributed lag functions performs similarly to MIDAS with unrestricted lag polynomials. In most of the cases, the simplest MIDAS with one lag of the factors is the best performing. Autoregressive dynamics plays a minor role. As far as the choice of the factor estimation technique is concerned, there are not significant differences among the different estimation methods. All Factor-MIDAS nowcasts can improve over quarterly factor forecast based on time-aggregated data, while the results compared to the state-space approach are less clear-cut and depend on the forecast horizon.

Kuzin et al. (2013) discuss nowcast pooling versus nowcasting of GDP growth for several countries, with single models in the presence of model uncertainty, with mixed-frequency and ragged-edge data. The now-

casts are based on MIDAS regressions with few indicators and Factor-MIDAS based on large datasets. The authors compare the performance of many alternative specifications with respect to alternative estimation methods, number of factors, indicators selected for MIDAS, the role of autoregressive dynamics. In their empirical analysis, they show that indicator models tend to outperform factor models in this ex-post evaluation. It is much more difficult to beat the benchmark when the models are selected based on information criteria or past performance. As a method to avoid the specification search, all the nowcasts and forecasts can be pooled together, using different selection schemes. This approach yields additional gains with respect to the factor specification based on past performance, in particular when all single-indicator and all Factor-MIDAS forecasts are combined together using inverse MSE weights, providing full support in favour of pooling for nowcasting and short-term forecasting.

9.6 Conclusions

In this paper we reviewed the literature concerning estimation and forecasting with mixed-sampling frequency and ragged-edge data. At the moment, temporal aggregation is still the predominant technique in the empirical applications: all data are sampled at the same lower-frequency. In filtering the data so that the variables have all the same frequency, potentially useful information is discarded. Empirical studies show that mixed-frequency data matter, the use of the procedures that allow taking different frequencies and the timeliness of the data into account improve the forecasts.

One of the early approaches to deal with mixed-sampling frequency is bridge equations, still very common in central banks, where a dynamic equation is estimated between the low-frequency variable and time-aggregated indicators. Separate high-frequency models provide forecasts of the high-frequency indicators, and these forecasts are then aggregated and plugged into the bridge equation. Empirical studies with bridge models show that the exploitation of intra-period information reduces the forecasting error in the majority of the cases. Bridge equations are a useful instrument especially in nowcasting, since the more information becomes available, the more accurate they are in terms of RMSE.

A second strand of research is based on mixed frequency regressions, where a low-frequency variable is explained by high-frequency indicators using parsimonious distributed lag models. MIDAS models are in general restricted to a limited set of variables, and estimated via NLS. Different weighting functions have been used in the literature, but which one is better to use is not clear-cut and depends on the specific analysis. Initially MIDAS models were applied to financial data, investigating stock market returns or future volatility, but recently MIDAS regressions have been employed to forecast macroeconomic variables, providing promising results for short-term forecasting.

Another, line of research relies on the state-space framework, in connection with both factors and VAR representations. The state-space setup treats the low-frequency variable as a high-frequency series with missing observations. The use of the Kalman filter allows real-time filtering, i.e., taking quarterly economic activity explained by monthly indicators as example, it is possible to obtain an estimate of GDP growth in each month. However, because of the intensive computation required by this framework and a relatively high number of parameters to be estimated, only models with few variables can be implemented. On the other hand, also structural applications can be considered, see e.g. Foroni and Marcellino (2014b), Foroni and Marcellino (2014c) and Marcellino and Sivec (2014). Within the class of factor models, different factor estimation techniques are described in the literature to handle the ragged-edge problem.

Recently, mixed frequency factor models and MIDAS regressions have been merged into Factor-MIDAS, which allows to forecast a low-frequency variable, exploiting the monthly information available in large datasets and handling unbalanced data as typical in real-time. Various factor estimation methods have been employed, without significant differences in their forecasting performance. Evidence shows that taking into account higher frequency information and exploiting the most recent observations pays off: Factor-MIDAS outperform quarterly factor forecasts based on time-aggregated data.

In summary, there is consensus that exploiting data at different frequencies matters, but it is not clear which method is superior. State-space models are a system approach and allow the estimation of the missing high-frequency data thanks to the use of the Kalman filter. Within this class of models, different ways to estimate VAR parameters or factors taking into account the unbalancedness of the datasets have been proposed, but the differences don't seem to be so pronounced. At the same time, the state-space approaches generally work only if the number of variables in the model is quite low, due to a dramatic increase in the number of parameters and associated complexity of the estimation. MIDAS models appear to be more robust to mis-specification compared both to bridge equation models and state-space approaches, and computationally simpler.

Bibliography

Aastveit K., Karsten R. and Jore A.S. (2011). Nowcasting GDP in Real-Time: A Density Combination Approach, mimeo, Norges Bank.

Altissimo F., Cristadoro R., Forni M., Lippi M. and Veronese G. (2010). New Eurocoin: Tracking economic growth in real time, The Review of Economics and Statistics, 92(4), 1024-1034.

Andreou E., Ghysels E. and Kourtellos A. (2010). Should macroeconomic forecasters use daily financial data and how?, University of Cyprus WP, 09-2010.

Angelini E., Henry J. and Marcellino M. (2006). Interpolation and backdating with a large information set, Journal of Economic Dynamics and Control, 30(12), 2693-2724.

Angelini E., Camba-Méndez G., Giannone D., Rünstler G. and Reichlin L. (2008). Short term forecasts of Euro Area GDP growth, Working Paper Series 949, European Central Bank.

Baffigi A., Golinelli R. and Parigi G. (2004). Bridge models to forecast the Euro area GDP, International Journal of Forecasting, 20(3), 447-460.

Bai J., Ghysels E. and Wright J. (2011). State space models and MIDAS regressions, Econometric Reviews, forthcoming.

Banbura M., Giannone D. and 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. and Reichlin L. (2012). Now-casting and the real time data flow, CEPR WP 9112.

Banbura M. and Modugno M. (2010). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data, Working Paper Series 1189, European Central Bank.

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.

Barhoumi K., Benk S., Cristadoro R., Den Reijer A., Jakaitiene A., Jelonek P., Rua A., Rünstler G., Ruth K. and Van Nieuwenhuyze C. (2008). Short-term forecasting of GDP using large monthly datasets. A pseudo real-time forecast evaluation exercise, Occasional Paper Series 84, European Central Bank.

Barhoumi K., Darné O., Ferrara L. and Pluyaud B. (2011). Monthly GDP forecasting using bridge models: application for the French economy, Bulletin of Economic Research, forthcoming.

Bencivelli L., Marcellino M. and Moretti G.L. (2012). Selecting predictors by Bayesian model averaging in bridge models, Banca d'Italia WP.

Breitung J. and Schumacher C. (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.

Camacho M. and Perez-Quiros G. (2010). Introducing the EURO-STING: Short Term INdicator of Euro Area Growth, Journal of Applied Econometrics, 25(4), 663-694.

Carriero A., Clark T. and Marcellino M. (2013). Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility, CEPR WP 9312.

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.

Diron M. (2008). Short-term forecasts of Euro Area real GDP growth. An assessment of real-time performance based on vintage data, Journal of Forecasting, 27(5), 371-390.

Doz C., Giannone D. and Reichlin L. (2011). A Two-step estimator for large approximate dynamic factor models based on Kalman filtering, Journal of Econometrics, 164(1), 188-205.

Chiu C., Eraker B., Foerster A., Kim T. and Seoane H. (2011). Bayesian Mixed Frequency VAR's, mimeo.

Evans N.D. (2005). Where Are We Now? Real-Time Estimates of the Macroeconomy, International Journal of Central Banking, 1(2).

Forni M., Hallin M., Lippi M. and Reichlin L. (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting, Journal of the American Statistical Association, 100, 830-840.

Foroni C. and Marcellino M. (2014a). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates, International Journal of Forecasting, 30, 554-568.

Foroni C. and Marcellino M. (2014b). Mixed-Frequency Structural Models: Identification, Estimation, and Policy Analysis, Journal of Applied Econometrics, forthcoming

Foroni C. and Marcellino M. (2014c). Mixed frequency structural VARs, Norges Bank WP 2014/01.

Foroni C., Guerin P. and Marcellino M. (2014). Markov-Switching Mixed Frequency VAR Models, International Journal of Forecasting, forthcoming.

Foroni C., Marcellino M. and Schumacher C. (2014). U-MIDAS: MIDAS regressions with unrestricted lag polynomial, Journal of the Royal Statistical Society, Series A, forthcoming

Forsberg L. and Ghysels E. (2007). Why do absolute returns predict volatility so well?, Journal of Financial Econometrics, 5(1), 31-67.

Frale C., Marcellino M., Mazzi G.L. and Proietti T. (2010). Survey data as coincident or leading indicators, Journal of Forecasting, 29(1-2), 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, 439-470.

Galvao A.B. (2007). Changes in Predictive Ability with Mixed Frequency Data, Queen Mary University WP, 595.

Ghysels E. (2011). Mixed frequency vector autoregressive models and the consequences of ignoring high frequency data, mimeo.

Ghysels E., Rubia A. and Valkanov R. (2009). Multi-period forecasts of volatility: direct, iterated and mixed-data approaches. University of North Carolina WP.

Ghysels E., Santa-Clara P. and Valkanov R. (2004). The MIDAS touch: MIxed DAta Sampling regression models, mimeo, Chapel Hill, N.C.

Ghysels E., Santa-Clara P. and 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. and Valkanov R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies, Journal of Econometrics, 131(1-2), 59-95.

Ghysels E., Sinko A. and Valkanov R. (2007). MIDAS regressions: Further results and new directions, Econometric Reviews, 26(1), 53-90.

Ghysels E. and Wright J. (2009). Forecasting professional forecasters, Journal of Business and Economic Statistics, 27(4), 504-516.

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.

Guerin P. and Marcellino M. (2013). Markov Switching MIDAS models, Journal of Business and Economic Statistics, 31(1), 45-56.

Harvey A.C. (1985). Applications of the Kalman filter in econometrics, in Bewley T.F. (ed.) Advances in Econometrics: Volume 1: Fifth World Congress.

Harvey A.C. and Pierse R.G. (1984). Estimating Missing Observations in Economic Time Series, Journal of the American Statistical Association, Volume 79, 125-131.

Harvey A.C. (1989). Forecasting structural time series models and the Kalman filter, Cambridge University Press.

Harvey A. and Chung C.H. (2000). Estimating the Underlying Change in Unemployment in the UK, Journal of the Royal Statistical Society, Series A, Vol. 163, Part 3, 303-339.

Ingenito R. and Trehan B. (1996). Using monthly data to predict quarterly output, Economic Review, Federal Reserve Bank of San Francisco, pages 3-11.

Koenig E., Dolmas S. and Piger J. (2003). The use and abuse of real-time data in economic forecasting, Review of Economics and Statistics, 85(3), 618-628.

Kuzin V., Marcellino M. and Schumacher C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area, International Journal of Forecasting, 27, 529-542.

Kuzin V., Marcellino M. and Schumacher C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries, Journal of Applied Econometrics, 28(3), 392-411.

Lanning S. (1986). Missing observations: A simultaneous approach versus interpolation by related series, Journal of Economic and Social Measurement, 14(1), 155-163.

Marcellino M. (1998). Temporal disaggregation, missing observations, outliers, and forecasting: a unified non-model based procedure, Advances in Econometrics, 13, 181-202.

Marcellino M., Porqueddu M. and Venditti F. (2013). Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility, Banca d'Italia, Temi di Discussione, 896.

Marcellino M. and 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.

Marcellino M. and Sivec V. (2014). Monetary, Fiscal and Oil Shocks: Evidence based on Mixed Frequency Structural FAVARs, mimeo.

Mariano R. and Murasawa Y. (2003). A new coincident index of business cycles based on monthly and quarterly series, Journal of Applied Econometrics, 18(4), 427-443.

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.

Mittnik S. and Zadrozny P.A. (2005). Forecasting quarterly German GDP at monthly intervals using monthly Ifo business conditions data, in Ifo survey data in business cycle and monetary policy analysis, ed. by J.-E. Sturm, and T. WollmershÄauser, pp. 19-48. Heidelberg: Physica-Verlag.

Monteforte M. and Moretti G. (2010). Real-time forecasts of inflation: the role of financial variables, Banca D'Italia WP, 767.

Schorfheide F. and Song D. (2011). Real-time forecasting with a mixed frequency VAR, mimeo.

Stark T. (2000). Does current-quarter information improve quarterly forecasts for the U.S. economy?, Working Papers 00-2, Federal Reserve Bank of Philadelphia.

Stock J. and Watson M. (1989). New indexes of coincident and leading indicators, NBER Macroeconomics Annual, 4, 351-394.

Stock J. and Watson M. (2002). Forecasting using principal components from a large number of predictors, Journal of the American Statistical Association, 97(460), 1167-1179.

Viefers P. (2011). Bayesian inference for the mixed-frequency VAR model, DIW Discussion Papers, 1172.

Wallis K. (1986). Forecasting with an econometric model: The 'ragged-edge' problem, Journal of Forecasting, 5, 1-13.

Wohlrabe K. (2009). Forecasting with Mixed-frequency Time Series Models, Ph.D. dissertation, University Munich.

Zadrozny P.A. (1988). Gaussian-Likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies, Econometric Theory, 4(1), 108-124.

IV Rapid Estimates based on Factor Models

10 Dynamic Factor Models: A Review of the Literature

Authors

Karim Barhoumi
IMF
kbarhoumi@imf.org

Olivier Darné
Université de Nantes
olivier.darne@univ-nantes.fr

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

10.1 Introduction

In recent years, the volume of available financial and economic data has led econometricians to develop or adapt methods to efficiently summarize the information contained in these large data bases. In applied macroeconomics, it is frequently the practitioner who has the tricky task of identifying, from among the large number N of variables available to him, the few variables of interest that will enable him to best solve his problem.

For example, economic growth and inflation forecasts are made in national and international institutions that have access to large volumes of data from surveys of households and businesses and various series on prices and real activity, such as the industrial production index (IPI), household consumption, the unemployment rate, etc. Similarly, central banks conduct monetary policy in a data-rich environment, looking at macroeconomic activity and the different financial markets on a regular basis and tracking numerous monetary aggregates.

A number of econometric methods have been proposed in the literature for working in such data-rich environments. For example, to explain the changes in a particular variable using a vast set of exogenous variables N in a linear regression model, the so-called *general-to-specific* (see Krolzig and Hendry (2001)) proposes an algorithm that makes it possible to select just a few variables from among those variables N. Similarly, vector autoregressive (VAR) models are recognized as allowing for simultaneous modeling of variables in a multivariate context. Traditionally, VAR models use a small number of variables to avoid inflating the number of parameters to be estimated. To remedy this problem, Bayesian approaches have been proposed to estimate VAR models with a high number of variables N by imposing restrictions (see, for example, De Mol et al. (2008)). Finally, if we consider the problem of predicting a particular variable when we have a large number of variables N that are potentially very relevant, we can imagine estimating N linear regressions, which then provide N forecasts that we will seek to combine (see, for example, Newbold and Harvey, (2002), for forecast combination methods). Eklund and Kapetanios (2008) also provide a review of the literature on the various forecasting techniques using large data sets.

Among the different methodologies proposed in the literature, dynamic factor models have grown significantly in popularity since the early 2000s and have been shown to be very useful in macroeconomic analysis and forecasting in a data-rich environment. These models can be used to summarize the information contained in a large number of economic variables into a small number of factors common to the set of variables. In this type of model, the N variables (x_{it}), for $i = 1, \ldots, N$ and $t = 1, \ldots, T$,where t refers to the time index, are each assumed to be the sum of two unobservable orthogonal components: one component resulting from the factors that are common to the set of variables, (χ_{it}), and an idiosyncratic component (ξ_{it}). The component (χ_{it}) is obtained by extracting a small number $r \geq 1$ of common factors (F_{jt}), $j = 1, \ldots, r$ from all of the variables present in the data set. Often, by extension, this component (χ_{it}) is identified by the term "common component" which we will also use in this article. The idiosyncratic component (ξ_{it}) covers the shocks specific to each of the variables. Thus, in a factor model of dimension $(N \times 1)$, each element of the vector $X_t = (x_{1t}, ..., x_{Nt})'$, assumed to be zero mean, can be written as follows:

$$x_{it} = \chi_{it} + \xi_{it},$$

or :

$$x_{it} = \lambda_{i1} F_{1t} + ... + \lambda_{ir} F_{rt} + \xi_{it}$$

For $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The loadings (λ_{ij}), for $i = 1, \ldots, N$ et $j = 1, \ldots, r$, represent the contributions of the variable i to the common factor (F_t) of dimension $(r \times 1)$ such as $F_t = (F_{1t}, ..., F_{rt})'$. The vector $(\xi_t) = (\xi_{1t}, ..., \xi_{Nt})'$ of dimension $(N \times 1)$ is a vector consisting of N idiosyncratic components. The vectorial form of the model is presented as follows, for all $t = 1, \ldots, T$:

$$X_t = \Lambda F_t + \xi_t, \tag{10.1}$$

where Λ is the weighting matrix of dimension $(N \times r)$. The matrix version is given as:

$$X = F\Lambda' + \xi, \tag{10.2}$$

where X is of dimension $(T \times N)$, F is of dimension $(T \times r)$, Λ is of dimension $(N \times r)$ and ξ is of dimension $(T \times N)$.

Given the rapid development of dynamic factor models in applied macroeconomics, we felt that the time was right to propose a review of the literature on these models so as to recap the current situation for practitioners. In this article, we begin by presenting the so-called traditional or classical factor models, which were developed initially for a small number of variables with common movements. We distinguish between static and dynamic approaches for these models. Then, we describe approximate factor models, which can take a large number of variables into account, again in a static or dynamic context. Next, we present some estimation models proposed in the literature. A crucial aspect of these models is the selection of the number of common factors r to use in the analysis and, so, we provide a review of the various information criteria developed to select the optimal number of factors. There are many applications of factor models in the empirical economic literature, including, for example, asset pricing models (Ross (1976)), consumer theory (Gorman (1981); Lewbel (1991)), performance assessment and risk measurement in finance (Campbell et al. (1997)). In the final section, we focus on some recent applications that underscore the interest of this approach for macroeconomists, particularly (i) for the construction of short-term economic indicators, (ii) for macroeconomic forecasting, and (iii) for international macroeconomics and monetary policy analysis.

10.2 Factor models for a small number of variables (small N)

In this section, we present factor models used to model a small number of variables N, where, in practice, N is generally lower than 6 or 7 variables. We begin with the simplest non-dynamic models (static factors) and, then, look at dynamic models, ending with a few recent extensions of this type of model.

10.2.1 Static factor models (SFM)

In this type of model, a small number of unobservable variables r provides a linear explanation of a small number of observed variables N so that $r < N$. In the applications presented in the final section of our article, the number of variables is such that $N \leq 7$ and a single factor can generally explain most of the variance, i.e. $r = 1$. The series are assumed to be stationary, to have finite variance, and to be standardized. We put forward the following hypotheses, which could subsequently be abandoned:

(SH1) The factors (F_t) are centered, $E(F_t) = 0$, and are mutually orthogonal for all t, *i.e.* : $\forall t, E(F_{jt}F_{j't}) = 0$ for $j \neq j'$. Consequently, the variance-covariance matrix of (F_t), $\Sigma_F = E(F_t F_t')$, is a diagonal matrix.

(SH2) The idiosyncratic processes (ξ_{it}) et $(\xi_{i't})$ are mutually orthogonal for all $i \neq i'$, with $E(\xi_t) = 0$. Consequently the variance-covariance matrix (ξ_t) is a diagonal matrix: $\Sigma_\xi = E(\xi_t \xi_t') = diag(\sigma_1^2, ..., \sigma_N^2)$;

(SH3) The factors (F_t) and idiosyncratic noise $(\xi_{it})_{i=1,...,N}$ are not correlated, *i.e.* : $\forall i, j, t, t'$ we have: $E(F_{jt}\xi_{it'}) = 0$;

(SH4) The variables are assumed to be independent and identically distributed over time (the so-called IID hypothesis), so that, in particular, for $t \neq t'$, $E(F_{jt}F_{jt'}) = 0$ and $E(\xi_{it}\xi_{it'}) = 0$.

The model given by equation (10.1) represents the static factor model (SFM) in which the factors (F_t) do not possess their own dynamic and the relationship between the factors and variables is linear with constant weights over time. This model can be estimated either by assuming that the variables are IID (hypothesis SH4), or by assuming that there is a time dynamic within the variables (SH4 is abandoned).

Assuming that (F_t) and (ξ_t) are not correlated and are zero mean, then the variance-covariance matrix for the static factor model, denoted $\Sigma_X = E(X_t X_t')$, is given by:

$$\Sigma_X = \Lambda \Sigma_F \Lambda' + \Sigma_\xi \tag{10.3}$$

By normalizing the variance-covariance matrices of (F_t), $\Sigma_F = I_r$, and by assuming that the diagonal elements of the variance-covariance matrix Σ_ξ of (ξ_t) are bounded, we obtain:

$$\Sigma_X = \Lambda \Lambda' + \Sigma_\xi \tag{10.4}$$

For additional details, we refer to Lawley and Maxwell (1971) and Anderson (1984). The static factor model can, thus, be identified and estimated. The factorial analysis method is used for the static estimation of the factors. The weighting matrix Λ can be estimated by minimizing the sum of the squared residuals as follows:

$$\sum_{t=1}^{T} (X_t - \Lambda F_t)' (X_t - \Lambda F_t) \tag{10.5}$$

subject to the constraint $\Lambda' \Lambda = I_r$.

In this context, Doz and Lenglart (1999) establish the asymptotic properties of the estimator. Specifically, they show that this method produces convergent estimators even when the data used are autocorrelated, as is the case with time series. Moreover, they show empirically that this method provides a very good approximation of the dynamic method, while being easier to establish, which is an essential quality for forecasters who regularly estimate the model.

10.2.2 Exact or strict dynamic factor models (DFM)

Static factor models (SFM) are different from exact or strict dynamic factor models (DFM) in the sense that the latter incorporate a time dynamic. Thus, in the DFM, the common component can be seen as a sum of common shocks, whether contemporaneous or lagged. The model is, then, defined as follows:

$$x_{it} = \chi_{it} + \xi_{it}, \tag{10.6}$$

where:

$$\begin{aligned} \chi_{it} &= b_{i1}^0 u_{1t} + \ldots + b_{i1}^s u_{1,t-s} + b_{i2}^0 u_{2t} + \ldots + b_{i2}^s u_{2,t-s} + \ldots \\ &+ b_{iq}^0 u_{qt} + \ldots + b_{iq}^s u_{q,t-s}, \end{aligned} \tag{10.7}$$

where (u_t), of dimension $(q \times 1)$, is the vector of common shocks such as $u_t = (u_{1t}, u_{2t}, \ldots, u_{qt})'$, with $q \leq N$, and where s is the number of lags included in the model. The parameters (b_{il}^τ), for $\tau = 0, \ldots, s$, $i = 1, \ldots, N$ et $l = 1, \ldots, q$, represent the weights of the finite dynamic factors s. We speak of a *restricted* DFM when s is finite and a *generalized* DFM when s is infinite.[1]

[1] For a discussion of the relationship between restricted DFMs and generalized DFMs, see Giannone et al. (2006) or Forni et al. (2009).

Equation 10.7 can be rewritten as follows:

$$\chi_{it} = \sum_{l=1}^{q} b_{il}(L)u_{lt}, \tag{10.8}$$

where $b_{il}(z) = b_{il}^0 + b_{il}^1.z + \cdots + b_{il}^s.z^s$ are polynomials of degree s and where L is the lag operator so that, for all s, $L^s u_t = u_{t-s}$.

In a matrix form, equation 10.7 can be rewritten as:

$$\chi_{it} = B_i(L)u_t, \tag{10.9}$$

where $B_i(L) = (b_{i1}(L), \ldots, b_{iq}(L))$ is a q vector of polynomials with degree s.

Moreover, we assume (Bai and Ng (2007)) that the vector $B_i(L)$ can be decomposed as follows:

$$B_i(L) = \lambda_i^*(L)C(L), \tag{10.10}$$

where $\lambda_i^*(L)$ is a r vector of polynomials with degree s and where $C(L)$ is a matrix of dimension $(r \times q)$. Using equations 10.9 and 10.10, we can then write the common component in the following manner:

$$\chi_{it} = \lambda_i^*(L)F_t^*, \tag{10.11}$$

where $F_t^* = C(L)u_t$ is a vector of dimension r that refers to the *static factors*, and the common shocks u_t of dimension $(q \times 1)$ to the *dynamic factors*. A model with q dynamic factors can, thus, be considered as a model with $r = q(s+1)$ static factors.

In the context of small-dimension dynamic factor models, the estimation is generally done in the time domain by likelihood maximization, as proposed by Dempster et al. (1977), Shumway and Stoffer (1982), Watson and Engle (1983), and Stock and Watson (1989) [2].

To estimate the DFM when N is small, the following hypotheses are generally put forward:

(DH1) The factors (F_{jt}) and $(F_{j't})$ are mutually orthogonal, but the factors (F_{jt}) can be autocorrelated and are variance-covariance stationary,*i.e.* : $\forall j \neq j', \tau \neq 0$, $E(F_{j,t}) = 0$, $cov(F_{j,t}, F_{j',t-\tau}) = 0$ and $cov(F_{j,t}, F_{j,t-\tau})$ depends only on τ.

(DH2) The idiosyncratic processes (ξ_{it}) and $(\xi_{i't})$ are mutually orthogonal, but the processes (ξ_{it}) can be autocorrelated and covariance-stationary, *i.e.* : $\forall i \neq i', \tau \neq 0$, $E(\xi_{it}) = 0$, $cov(\xi_{i,t}, \xi_{i',t-\tau}) = 0$ and $cov(\xi_{i,t}, \xi_{i,t-\tau})$ depends only on τ.

(DH3) The factors (F_{jt}) and the idiosyncratic processes (ξ_{it}) are orthogonal for all i, j.

Based on these hypotheses, we can, then, attempt to estimate a dynamic factor model by likelihood maximization in the time domain, with the additional hypothesis of Normality for the model residuals. The maximum likelihood estimator is calculated by, first, placing the model in a space-state form and, then, using a Kalman-type recursive filter.

[2] Another method of estimating this type of model has been proposed by Sargent and Sims (1977) and Geweke (1977) in the frequency domain, based on a spectral analysis. We come back to this type of frequency domain estimation in section 10.4.

The DFM can be written in a space-state form, assuming that the common factors follow a VAR process of order p such as :

$$\Phi(L)F_t = \varepsilon_t \qquad \Leftrightarrow \qquad F_t = \sum_{\tau=1}^{p} \Phi_\tau F_{t-\tau} + \varepsilon_t, \tag{10.12}$$

and, for a given index i, the idiosyncratic process (ξ_{it}) follows an AR process of order p' in the following form:

$$\psi_i(L)\xi_{it} = \eta_{it} \qquad \Leftrightarrow \qquad \xi_{it} = \sum_{\tau=1}^{p'} \psi_{i\tau}\xi_{i,t-\tau} + \eta_{it}, \tag{10.13}$$

where (ε_t) and (η_{it}) are innovations of (F_t) and (ξ_{it}), respectively, so that (ε_t) and (η_{it}) are independent. $\Phi(.)$ and $\psi(.)$ are polynomials of order p and p', respectively, so that: $\phi(L) = I - \phi_1 L - \ldots - \phi_p L^p$ and $\psi_i(L) = I - \psi_{i1}L - \ldots - \psi_{ip'}L^{p'}$. The hypothesis of Normality is put forward for (ε_t) and (η_{it}). In practice, the orders p and p' of the lag polynomials must be selected prior to the estimation stage. This selection is generally done by minimizing an AIC-type information criterion (Akaike information criterion) or a BIC-type information criterion (Bayesian information criterion) or by using the Doz and Lenglart (1999) test. In empirical studies, $p = 2$ and $p' = 1$ are often shown to be sufficient to whiten the residuals.

This type of model shown by equations (10.1),(10.12) and (10.13) allows a space-state representation as follows:

$$X_t = c_t\beta_t + m_t Z_t + w_t, \tag{10.14}$$

where (Z_t) is a vector of n explanatory variables, for example, the lagged values of the observed variables (X_t), and where:

$$\beta_t = a_t\beta_{t-1} + v_t. \tag{10.15}$$

Equation (10.14) is the measure equation, which describes the relations between the unobservable states, of dimension r, , and the observable variables, of dimension n, where β_t represents the state vector:

$$\beta_t = \begin{bmatrix} F_t \\ \vdots \\ F_{t-p+1} \\ \xi_t \\ \vdots \\ \xi_{t-q+1} \end{bmatrix}$$

Equation (10.15) represents the state or transition equation, which describes the development of unobservable states. We see that a_t, c_t and m_t are matrices that can depend on time, of the dimensions $((p \times r + q \times N) \times (p \times r + q \times N))$, $(N \times (p \times r + q \times N))$ and $(N \times n)$, respectively, and where v_t is a Gaussian white noise vector of dimension $(p \times r + q \times N)$, w_t is a Gaussian white noise vector of dimension N, of the variance-covariance matrices Q_t and R_t respectively. In practice, the system is generally assumed to be invariant over time, *i.e.* a_t, c_t and m_t are constant. It is also assumed that for all t, $t' \neq t$, $E(v_t w_t') = 0$.

The model in its space-state form can then be estimated by maximum likelihood using a filtering method such as the Kalman filter. We refer, for example, to Hamilton (1994) for a description of the filtering algorithm. The maximum likelihood estimation algorithm can take a great deal of time as it requires inversion of a large dimensional matrix, even when N is small. In general, in the case of numerical optimization, the expectation-maximization (EM) algorithm is used, as proposed by Dempster et al. (1977) or Shumway and Stoffer

(1982)[3].

10.2.3 Recent extensions of factor models with small N

Several extensions of factor models with a small number of variables have recently been proposed to take certain data characteristics into account. We present two of these extensions below: Markov regime-switching models and mixed frequency models.

Markov regime-switching models

These models are directly linked to the Markov regime-switching processes introduced by Hamilton (1989) and assume that the common unobservable factors have their own dynamics governed by a two-regime Markov chain, denoted (S_t), with for all t, $S_t \in \{1, 2\}$. The idea of these models is to assume that the factors are related to the state of the economy, which itself evolves cyclically but non-periodically based on two economic phases that follow one another. We, then, assume that, for example, the first regime ($S_t = 1$) corresponds to the low phase of the business cycle and the second regime ($S_t = 2$) to the high phase of the business cycle. The model can easily be extended to a larger number of regimes, but the estimation problems then become tricky in that the model contains two latency levels, i.e., the common factors and the Markov chain.

An initial model of this type was proposed by Diebold and Rudebusch (1996), but the theoretical and empirical aspects were more broadly considered by Kim and Yoo (1995) and Kim and Nelson (1998). At the same time, Chauvet (1998) independently proposed a similar model. For example, in the case of a single factor (*i.e.* $r = 1$), for N centered stationary variables, the Markov regime-switching model can be defined as follows for $i = 1, \ldots, N$ et $t = 1, \ldots, T$:

$$x_{it} = \lambda_i' F_t + \xi_{it}, \tag{10.16}$$

with:

$$\phi(L) F_t = \mu(S_t) + \epsilon_t, \tag{10.17}$$

where the $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{ir})$, are the loadings, where, for each i, $(\xi_{it})_t$ follows a Gaussian autoregressive process of order one (AR(1)) of finite variance σ_i^2, where $(\epsilon_t)_t$ is a Gaussian white noise of unit variance and where $\phi(L) = I - \phi_1 L - \ldots - \phi_p L^p$. If we assume that $(S_t)_t$ is a first-order two-regime Markov chain, this means that the probability of S_t belonging to a regime at date t depends only on the probability of being in a certain regime at date $t - 1$, or:

$$P(S_t | S_{t-1}, S_{t-2}, S_{t-3}, \ldots) = P(S_t | S_{t-1}). \tag{10.18}$$

The transition probabilities $p_{12} = P(S_t = 2 | S_{t-1} = 1)$ and $p_{21} = P(S_t = 1 | S_{t-1} = 2)$ measure the probability of switching from one regime to the other. Similarly, the probabilities $p_{11} = 1 - p_{12}$ and $p_{22} = 1 - p_{21}$ measure the probability of remaining in the same regime, thus reflecting the degree of persistence of each regime. The estimation stage allows for the estimation for each date t of the expected, filtered and smoothed probabilities of being in a particular regime, respectively given by $P(S_t | \hat{\theta}, X_{t-1}, \ldots, X_1)$, $P(S_t | \hat{\theta}, X_t, \ldots, X_1)$ and $P(S_t | \hat{\theta}, X_T, \ldots, X_1)$, where $\hat{\theta}$ represents the set of estimated parameters of the model, which includes the N autoregressive parameters of the AR(1)models, the N idiosyncratic variances and the p parameters of the polynomial $\phi(.)$.

The parameters of this model can be estimated simultaneously by maximum likelihood, as proposed by Kim and Nelson (1998), or in two steps, by first estimating the common factor (F_t) in the time or spectral domain

[3] Alternatively, the Fisher scoring algorithm is used by Watson and Engle (1983).

(see above), and then adjusting an autoregressive regime-switching process on the estimated factor (see, on this point, Diebold and Rudebusch (1996)). From a theoretical standpoint, simultaneous estimation is preferable but, empirically, the maximization algorithm is often difficult to converge, in particular if the variables are very volatile. The two-stage estimation is more practical (see Darné and Ferrara (2011), for an application), but the second equation (10.17) then includes an error in the measure of the estimated factors that is not explicitly integrated in the model, which can thus create statistical inference problems.

Mixed frequency models

Numerous macroeconomic series are available to forecasters but do not necessarily have the same sampling frequency (or periodicity). The national accounts in particular, which most economists seek to predict, are available only on a quarterly basis while many economic indicators, such as the industrial production index (IPI), consumer expenditure by households, or opinion surveys are monthly. To be able to simultaneously handle these two periodicities in a single model, Mariano and Murasawa (2003) proposed a dynamic factor model in a space-state form that considers quarterly series as monthly series with missing values.

The aim of this type of model is to estimate a common factor with N variables, variables, some of which are quarterly and some of which are monthly. Thus, we write $(Y_{1,t})$ a vector of N_1 quarterly variables that are observable only in the third month t of the quarter and $(Y_{2,t})$ a vector of N_2 monthly variables, so that $N_1+N_2=N$. We assume here that these series (in logarithms) are integrated of order one. We also assume that there is a vector $(Y^*_{1,t})_t$ of N_1 unobservable monthly variables so that for all t, $Y_{1,t}$ is the geometric mean of $Y^*_{1,t}$ over the three months in a given quarter, *i.e.* :

$$\log(Y_{1,t}) = \frac{1}{3}\left[\log(Y^*_{1,t}) + \log(Y^*_{1,t-1}) + \log(Y^*_{1,t-2})\right]. \tag{10.19}$$

We note that this identity (10.19) differs from the arithmetical mean generally used in the quarterly accounts but makes it possible to implement a linear space-state form, in contrast to the arithmetical identity, that requires a non-linear form.

Mariano and Murasawa (2003), then, show that:

$$y_{1,t} = \frac{1}{3}y^*_{1,t} + \frac{2}{3}y^*_{1,t-1} + y^*_{1,t-2} + \frac{2}{3}y^*_{1,t-3} + \frac{1}{3}y^*_{1,t-4} \tag{10.20}$$

where $y_{1,t} = Y_{1,t} - Y_{1,t-3}$ et $y^*_{1,t} = Y^*_{1,t} - Y^*_{1,t-1}$, $(y_{1,t})$ being observable in all three periods and $(y^*_{1,t})$ being unobservable.

Under all of the standard hypotheses of dynamics on factors and idiosyncratic errors, and the Normality of residuals (see Mariano and Murasawa (2003), p. 430), one can show that the model with a factor (F_t) is written in the following form for all t :

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_1(\frac{1}{3}F_t + \frac{2}{3}F_{t-1} + F_{t-2} + \frac{2}{3}F_{t-3} + \frac{1}{3}F_{t-4}) \\ a_2 F_t \end{pmatrix} \tag{10.21}$$

$$+ \begin{pmatrix} \frac{1}{3}u_{1,t} + \frac{2}{3}u_{1,t-1} + u_{1,t-2} + \frac{2}{3}u_{1,t-3} + \frac{1}{3}u_{1,t-4} \\ u_{2,t} \end{pmatrix} \tag{10.22}$$

Where $a = (a_1', a_2')'$ is the weight vector of dimension N, (F_t) is the scalar common factor, $u_t = (u'_{1,t}, u'_{2,t})'$ is the idiosyncratic component of dimension N and where $y_{2,t} = Y_{2,t} - Y_{2,t-1}$. The model is, then, put in a space-state form (see Mariano and Murasawa (2003), p. 431) and estimated by maximum likelihood with the help of a Kalman filter.

Other approaches have been proposed to manage data with different periodicities simultaneously in factor models. For example, Aruoba et al. (2009) also propose a factor model containing four variables of different periodicities (daily, weekly, monthly and quarterly) to estimate U.S. GDP with a high periodicity. This indicator is actually updated weekly by the Philadelphia Federal Reserve on its internet site. Similarly, Camacho and Perez-Quiros (2010) and Camacho and Perez-Quiros (2011) propose a factor model that simultaneously treats variables with different periodicities and regime-switching in factors to estimate GDP growth in the Eurozone and Spain. For French data, Cornec and Deperraz (2006) construct a composite activity indicator for services based on monthly and quarterly survey data. Cornec (2006) also develops an indicator based on a mixed frequency factor model (see the last section on applications). More generally, we should point out that this type of approach makes it possible to treat the problem of missing data in series in an econometric model.

10.3 Approximate factor models (large N)

Although the concept of factor models is attractive, the traditional approach presented in the previous section has a number of limitations that are both theoretical and practical in nature.

1. The number of variables (N) is often larger than the number of observations (T) in economic data series. Consequently, potentially important information is lost when a small number of variables must be selected to respect the constraint that N be small;
2. Asymptotic convergence of estimators is assured when T tends to infinity and N is fixed, but not when N also tends to infinity;
3. *I.I.D.* hypotheses and hypotheses on the diagonality of the variance-covariance matrix of the idiosyncratic component Σ_ξ, which prohibit the cross correlation, are often too strong for economic data. This can result in a risk of misspecification;
4. The maximum likelihood estimation (MLE) is generally considered unachievable for factor models of large dimensions because the number of parameters to be estimated is too large (Bai (2004); Bai and Ng (2002));
5. The traditional approach makes it possible to consistently estimate the coefficients of the weighting factors (λ_i) by MLE when T is large, but not the common factors (F_t), for which only the estimated value can be obtained (Steiger (1979)). Meanwhile, in most economic problems, it is these common factors that are of greatest interest since they represent the common shocks, the diffusion indices, etc., for example.

To respond to a number of these limitations, the idea of factor models was generalized to allow for the manipulation of less strict hypotheses on the variance-covariance matrix of the idiosyncratic components by proposing an approximate factor structure. Non-parametric estimators of common factors based on the principal components have been suggested (Forni et al. (2000) ; Stock and Watson (2002)), their asymptotic properties being known when N is large. New methodologies have, consequently, been proposed.

10.3.1 Approximate static factor models (SFM)

Chamberlain and Rothschild (1983) are the first to introduce the so-called "approximate" factor structure concept by abandoning the hypothesis that idiosyncratic disturbances are not mutually correlated within the so-called "strict" factor structure, i.e., allowing for idiosyncratic errors to be weakly correlated. This concept makes it possible to obtain a non-diagonal variance-covariance matrix $\Sigma_\xi = E(\xi_t \xi_t')$. Moreover, Chamberlain and Rothschild (1983) show that a principal components analysis (PCA) is equivalent to a factor analysis (or to a maximum likelihood analysis under the hypothesis of Normality of the idiosyncratic component (ξ_{it})) when N increases to infinity. However, they assume that the variance-covariance matrix of the population, Σ_X, of dimension $(N \times N)$ is known. Connor and Korajczyk (1986), Connor and Korajczyk (1988), Connor and

Korajczyk (1993) study the case of an unknown variance-covariance matrix Σ_X and suggest that, when N is larger than T, the factor model can be estimated by applying a PCA to the variance-covariance matrix Σ_X, of dimension $(T \times T)$.

Connor and Korajczyk (1986) establish the coherence of factors estimated by PCA when T is fixed and N tends to infinity in the context of approximate factor models but they provide no formal argument when N and T simultaneously tend to infinity.[4] Stock and Watson (1999) study the uniform coherence of estimated factors and derive convergence rates for large T and N. The convergence rate is also studied by Bai and Ng (2002). Finally, Bai (2004) shows that the PCA estimator of the common component is asymptotically Gaussian, converging to a rate equal to $\min(N^{1/2}, T^{1/2})$, even when the idiosyncratic component is serially correlated and/or heteroskedastic when N and T are large.[5]

10.3.2 Approximate dynamic factor models (DFM)

Forni and Lippi (1997), Forni and Reichlin (1998), Forni et al. (2000) and Forni et al. (2004) extend approximate factor models by considering dynamic factor models of large dimensions and introduce different methods for the estimation of this type of model. These models are referred to as "generalized" because they combine both dynamic and approximate structures, i.e., they generalize exact dynamic factor models by assuming that the number of variables N tends to infinity and by allowing idiosyncratic processes to be mutually correlated.

Forni et al. (2000) and Forni et al. (2004) expand the dynamic principal components analysis introduced by Brillinger (1981) when N is large. The estimation proposed by Brillinger (1981) generalizes the static PCA by placing the analysis in the frequency domain. First, the spectral density of the vector X_t is estimated using a consistent spectral density estimator, denoted $\widehat{S}(\omega)$, for a frequency $\omega \in]0, 2\pi]$. Then, the eigenvectors corresponding to the largest q eigenvalues of this spectral matrix are calculated. Finally, one returns to the time domain by applying the inverse Fourier transform to these eigenvectors, to recover the estimators of the time series in dynamic principal components (see next section).

Brillinger (1981) obtains distributional results when N is fixed and T tends to infinity. Forni et al. (2000) show that the dynamic PCA provides a consistent estimate of the common component when both N and T increase. Forni et al. (2004) discuss the coherence conditions and convergence rates.

It has been shown that the principal components are convergent estimators of factors, both in the static context (Bai and Ng (2002); Stock and Watson (2002); Forni et al. (2000) and Bai (2004) and in the dynamic context (Forni et al. (2000) and Forni et al. (2004)).

Approximate factor models have several advantages over strict models. They are flexible and appropriate under general hypotheses on measurement errors and, usually, on the cross-correlation of idiosyncratic components. The misspecification error resulting from the approximate structure of the idiosyncratic component disappears when N and T are large, as long as the cross-correlation of the idiosyncratic processes is relatively small and that of the common components increases across the transverse dimension when N increases. These conditions are introduced in Chamberlain and Rothschild (1983) and used, reinterpreted and expanded in Connor and Korajczyk (1986), Connor and Korajczyk (1988), Connor and Korajczyk (1993), Forni and Lippi (1997), Forni et al. (2000) and Stock and Watson (2002), respectively. In short, approximate factor models have two important advantages over traditional factor models:

[4] Ding and Hwang (1999) obtain results on the coherence for PCA estimation of traditional exact factor models when N and T tend to infinity.

[5] Jones (2001) and Boivin and Ng (2005) propose weighted PCA estimators by considering the problem of non-linear generalized least squares as follows: $\min_{F_1,\ldots,F_T,\Lambda} \sum_{t=1}^{T} (X_t - \Lambda F_t)' \Sigma_\xi^{-1} (X_t - \Lambda F_t)$, with (ξ_t) being *i.i.d* and of the normal distribution $N(0, \Sigma_\xi)$. Stock and Watson (2005a) extend the weighted PCA approach by assuming an autoregressive structure of weak order for (ξ_t).

1. The idiosyncratic components can both be weakly mutually correlated and show little heteroskedasticity. This can reflect the condition in which all the eigenvalues of the idiosyncratic variance-covariance matrix $\Sigma_\xi = E(\xi_t \xi_t^{'})$ are bounded. Thus, the absolute mean of the covariances is bounded, *i.e.* : $\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |E(\xi_{it}\xi_{jt})| < \infty$ (Stock and Watson (2002))[6] ;
2. In this type of model, it is possible to have a weak correlation between the factors (F_t) and the idiosyncratic components (ξ_t).

10.3.3 Recent extensions of approximate factor models

Among the recent extensions of dynamic factor models when N is large, we would like to mention FAVAR models, models whose parameters vary over time and mixed frequency models. The applications of these different types of models are presented in the last section on applications. We will see that these models assume that the variables are stationary. For the development of dynamic factor models on data that are non-stationary, we refer, for example, to Pĕna and Poncela, (2006a) and Pĕna and Poncela, (2006b). Moreover, Banerjee and Marcellino (2009) have extended the FAVAR models to the factor-augmented error correction model, which makes it possible to integrate variables that are non-stationary.

FAVAR models

To remedy the problem of missing variables, generally encountered in traditional VAR and SVAR (structural VAR) modeling, Bernanke et al. (2005) propose using factor-augmented VAR (FAVAR) models, particularly in the context of monetary policy analysis.

The FAVAR model can be described by the following equation:

$$X_t = \Lambda F_t + B X_{t-1} + \xi_t \tag{10.23}$$

where (X_t) represents the endogenous variables of a traditional VAR model, such as in Bernanke et al. (2005) and Boivin et al. (2009), (F_t) the common factor, Λ the weighting matrix, and (ξ_t) the idiosyncratic component. In Stock and Watson (2005a), B is a diagonal matrix $D(L)$ with a lag polynomial $\delta_i(L)$ on the i-th diagonal. It is also conceivable to specify a short-term dynamics, of the first-order autoregressive type, for example, on the common factor (F_t) and on the idiosyncratic component (ξ_t).

Stock and Watson (2005a) propose using an iterative procedure to estimate the FAVAR model given by equation (10.23). This procedure begins with an initial estimation of the static factor $\hat{F}_t$ using a PCA. Then, the weighting matrix $\hat{\Lambda}$ and the coefficients $\hat{B}$ are estimated by ordinary least squares. Finally, the $\hat{F}_t$ factors are re-estimated by the principal components of $X_t - \widehat{B} X_{t-1}$ and this procedure is iterated until convergence. Boivin et al. (2009) also use this iterative procedure with $B = 0$ as the initial estimation. Bernanke et al. (2005) propose estimating the unobservable factors in two steps: (1) the principal components of the informational variables are, first, calculated by ignoring the presence of the observable variables; (2) equation (10.23) is, then, estimated by integrating the factors estimated in the previous step. Other works on monetary policy analysis involve the estimation of factors using a dynamic factor model, these estimated factors then being introduced into a VAR model as additional regressors (see for examples, Giannone et al. (2004) and Favero et al. (2005)).

Time-varying parameter models

Some authors are also interested in factor models for which, for example, the weights grouped in the matrix Λ of equation (10.1), vary over time (see for example Motta et al. (2011)). This type of approach is promising since it makes it possible to integrate structural changes in respect of the source and amplitude of the shocks

[6] The precise technical condition allowing for a weak correlation of idiosyncratic terms varies from one study to another but, in general, this condition limits the contribution of the idiosyncratic covariances to the variance-covariance matrix of X_t when N is large.

into the modeling, as well as their channels of transmission to the economy. As well, this type of modeling incorporates non-linearity into relations. In particular, this makes it possible to assess whether behaviors change over the course of a business cycle.

In the area of FAVAR models, some recent papers incorporate a dynamic structure by allowing a change over time (i) of the weights, (ii)of the autoregressive dynamics of the factors, or (iii) of the variance of innovations. Thus, a possible specification of a time-varying FAVAR (TV-FAVAR) model is as follows:

$$X_t = \Lambda_t F_t + B_t X_{t-1} + \xi_t, \tag{10.24}$$

with $E(\xi_t) = 0$ and $E(\xi_t \xi_t') = \Sigma_{\xi_t}$.

Generally, the FAVAR model given by equation (10.24) is estimated in a Bayesian context, as is done, for example, in the articles of Del Negro et al. (2008), Mumtaz and Surico (2009) or Baumeister et al. (2010). Similarly, Kose et al. (2003) propose an estimation of dynamic factor models based on an MCMC (*Monte Carlo Markov-Chain* approach in a Bayesian context.[7]

Eickmeier et al. (2015) propose an alternative approach by developing a two-step standard estimation model by maximum likelihood. The first step consists in estimating the factors in a static context (see next section). Then, in a second step, the model with non-constant parameters is estimated equation by equation. Thus, each univariate regression equation is put in a space-state form, then estimated in the traditional fashion using a Kalman filter. This approach requires assuming that FAVAR model equations are conditionally independent.

Mixed frequency models

Finally, in the case of a large number N of available variables in the data set, variables with different periodicities can be managed using a MIDAS (MIxed DAta Sampling) regression proposed by Ghysels and his co-authors (see for example Ghysels et al. (2007) for a presentation). The MIDAS approach can be used to explain a sampled variable with a certain periodicity (such as annual or quarterly) by variables with a higher periodicity (such as monthly or daily) without having to, first, aggregate the higher periodicity data. This approach is based on a standard linear regression equation but involves the estimation of a weight function depending on a hyper-parameter with a smaller dimension compared with the initial dimension of the problem.

Let us assume, for example, that we are seeking to estimate the quarterly GDP growth rate of an economy, denoted (y_t), assumed to be stationary, for a number of quarters T, the index t here designating the quarter. Let us also assume that, with the help of one of the methods shown in the next section, we have estimated a single monthly stationary factor $(\hat{F}_t^{(m)})$ (*i.e.* : $r = 1$) based on a large monthly data set. We, thus, observe $m = 3$ times $(\hat{F}_t^{(m)})$ over the period $[t-1, t]$.

The standard MIDAS equation makes it possible to link the quarterly variable to be explained to the monthly estimated factor as follows:

$$y_t = c_0 + c_1 \, G(\theta) \, \hat{F}_t^{(m)} + \epsilon_t, \tag{10.25}$$

where c_0 and c_1 are parameters to be estimated and where (ϵ_t) is assumed to be a Gaussian white noise of finite variance that will also need to be estimated. The term $G(\theta)$ controls the polynomial weights, which allows the mixing of frequencies. In fact, the MIDAS specification consists in smoothing the past values of the

[7] MCMC methods are numerical methods that create long Markov chains, making it thus possible to obtain samples distributed asymptotically according to a certain distribution.

$(\hat{F}_t^{(m)})$ using the polynomial $G(\theta)$ of the following form:

$$G(\theta) = \sum_{k=1}^{K} g_k(\theta) L^{(k-1)/m} \tag{10.26}$$

where K is the number of points on which the smoothing operates, L is the lag operator, so that, for any monthly variable $x_t^{(m)}$, $L^{\tau/m} x_t^{(m)} = x_{t-\tau/m}^{(m)}$ and $g_K(.)$ is the weight function, which can take various forms. As in Ghysels *et al.* (2007), we generally use a two-parameter Almon polynomial $\theta = (\theta_1, \theta_2)$ so that :

$$g_k(\theta) \equiv g_k(\theta_1, \theta_2) = \frac{\exp\left(\theta_1 k + \theta_2 k^2\right)}{\sum_{k=1}^{K} \exp\left(\theta_1 k + \theta_2 k^2\right)} \tag{10.27}$$

The θ parameter is part of the estimation problem. It is influenced by the information contained in the last K values of $(\hat{F}_t^{(m)})$, the size of the window K being an exogenous parameter. Other specifications can be considered in the literature for equation (10.25), particularly by adding monthly explanatory variables or autoregressive terms for the target variable (y_t). Similarly, other weight functions can be considered.

In terms of application, Marcellino and Schumacher (2010) propose an approach in which they, first, estimate the monthly factors based on a data set of 111 representative variables of the German economy. They, then, use these factors to forecast the quarterly German GDP using a MIDAS regression (so-called factor-MIDAS approach). The authors show the usefulness of such an approach in making better use of the most recent data for short-term macroeconomic forecasting purposes. As for the MIDAS approach itself, some examples of applications, in forecasting, for example, are found in the articles by Clements and Galvao (2008) and Ferrara and Marsilli (2013).

10.4 Estimation of factor models for large N

In this section, we present the main estimation methods of factor models, whether static or dynamic, when the number of variables is high (large N). In this case, the usual methods based on maximizing likelihood run into the problem of the dimension of the parameter to be estimated.

10.4.1 Static factor models: the Stock and Watson (2002) approach

One of the first approximate factor models is the one proposed by Stock and Watson (2002), which is based on a static PCA. The PCA is used since it allows for the estimation of both the parameters and the factors of the model given by equation (10.1) by maximizing the variance explained by the initial variables, for a small number r of static factors (F_t). The main aim of the Stock and Watson (SW) approach is to approximate the factors by a linear combination of the data $\widehat{F}_{j,t} = \widehat{W}_j' X_t$, for $j = 1,...,r$, that maximizes the variance of the estimated factors $\widehat{W}_j' \widehat{\Sigma}_x \widehat{W}_j$, where $\widehat{\Sigma}_x = (1/T) \sum_{t=1}^{T} X_t X_t'$ is the empirical variance-covariance matrix of the vector of the initial standardized data X_t.

Under the following normalization assumption: $\widehat{W}_j' \widehat{W}_{j'} = 1$ for $j = j'$ and $\widehat{W}_j' \widehat{W}_{j'} = 0$ for $j \neq j'$, the maximization problem can, then, be transformed into the solution of a eigenvalues problem:

$$\widehat{\Sigma}_x \widehat{W}_j = \widehat{\mu}_j \widehat{W}_j, \tag{10.28}$$

where $\widehat{\mu}_j$ is the j-th eigenvalue and $\widehat{W}_j$ is the associated eigenvector of dimension $(N \times 1)$. Once they have been calculated, the highest N eigenvalues are classified in decreasing order. Then, the eigenvectors are, in turn, classified in decreasing order with respect to the highest r eigenvalues. The factors proposed by SW are, then, written as follows:

$$F_t^{SW} = \widehat{W}' X_t, \tag{10.29}$$

where $\widehat{W}$ is the matrix of dimension $(N \times r)$ of the stacked eigenvectors $\widehat{W} = (\widehat{W}_1, ..., \widehat{W}_r)$[8].

However, the Stock and Watson approach does not allow for use of the different dynamics that may exist between the variables used. To take account of this dynamic structure in factor models, several alternatives to the static factor model have been proposed in the literature. Specifically, there are two main types of dynamic factor models or approaches. Developed by Doz et al. (2011) and Doz et al. (2012), the first one is based on a space-state representation of models in the time domain. Proposed by Forni et al. (2004) and Forni et al. (2005), the second one is based on the spectral domain. We, now, present the estimation strategies of these different dynamic factor models.

10.4.2 Dynamic factor models

Time domain approach

Doz et al. (2011) and Doz et al. (2012) propose a dynamic factor model that can be represented in a space-state form. Specifically, they estimate their dynamic factor model using two different approaches. The first one is the so-called two-step approach (Doz et al. (2011)). The second one is based on the quasi maximum likelihood (Doz et al. (2012)).

According to Doz et al. (2011), for a number r of factors and q, of dynamic shocks, the estimation is carried out in two steps. In the first step:

1. $\widehat{F}_t$ is estimated using a PCA, as an initial estimate;
2. Then, equations (10.6) and (10.11) are estimated using the estimated factor from the previous step, $\widehat{F}_t$, to obtain both $\widehat{\lambda}_i^*(L)$ and the variance-covariance matrix of the residuals $\widehat{\xi}^*$, denoted $\widehat{\Sigma}_{\xi^*}$. To obtain an estimate of $C(L)$, appearing in equation (10.10), DGR (2011) apply a decomposition of eigenvalues to the matrix $\widehat{\Sigma}_{\xi^*}$ by taking into account the number of dynamic shocks q. Let us introduce the matrix M of dimension $(r \times q)$ corresponding to the largest q eigenvalues and the matrix P of dimension $(q \times q)$ containing the largest q eigenvalues in its diagonal and zeros elsewhere. The estimate of $C(L)$ is, then, obtained by $\widehat{C}(L) = M \times P^{-1/2}$.

In a second step, the coefficients and parameters of the system described by equations (10.6) and (10.11) are considered to be known and provided by the first step. The model is, then, written in a space-state form and the Kalman filter is applied to obtain new estimates of the factors.

In their alternative approach, Doz et al. (2012) estimate an approximate dynamic factor model using the quasi maximum likelihood method [9]. The main aim of this approach is to consider the strict factor model as a misspecification of the approximate factor model and to analyze the properties of the maximum likelihood indicator of the factors under this misspecification. This estimator is called the quasi maximum likelihood in the sense of White (1982). By analyzing the properties of the maximum likelihood estimator under several sources of misspecifications, such as an omitted serial correlation of the observations or a cross-sectional correlation of the idiosyncratic components, Doz et al. (2012) show that these misspecifications do not affect the robustness of the common factors, particularly for fairly large N and T. More specifically, this estimator is a valid parametric alternative for the estimator resulting from a PCA. The model defined by means of equations (10.6) and (10.11) can be put in a space-state form, with a number of states equal to the number of common factors r. It is noteworthy that the estimation of the parameters of the model, particularly the common factors, by the quasi maximum likelihood can be approximated by their anticipated values, using the Kalman filter[10].

[8] Stock and Watson (1998) develop theoretical results for this methodology.

[9] Jungbacker and Koopman (2008) propose new results for the estimation of a dynamic factor model using the maximum likelihood method and a Bayesian method based on Markov chains. Jungbacker et al. (2011) adapt this approach in the context of missing data.

[10] The likelihood can be maximized by means of the EM algorithm, which requires the use of the Kalman filter for each iteration.

These dynamic factor models have also been called restricted dynamic factor models, since the r static factors are caused by a number q of dynamic factors, with $q \leq r$ (Forni et al. (2005) ; Hallin and Liska (2007)).

Kapetanios and Marcellino (2004) also propose an approach based on a space-state representation. Their approach is based on the use of specific subspaces in which the factors are estimated. This subspace algorithm can be used to estimate factors without having to specify or identify the model entirely in its space-state form.

Frequency domain approach

In a series of articles, Forni, Hallin, Lippi and Reichlin (2000, 2003, 2004, 2005) (FHLR) propose a dynamic PCA in the frequency domain, also called a generalized dynamic factor model, to estimate dynamic factors[11]. The purpose of their model is to identify the dynamic structure of a factor model. The dynamic factor model is given by equations (10.6) and (10.7). The method proposed by FHLR makes it possible to estimate dynamic factors in a first step and, then, obtain the static factors from the estimated dynamic factors in a second step. The approach proposed by FHLR aims to estimate both the dynamic factors and their covariances. This estimation is performed to maximize the variance of the common component under certain orthogonality restrictions. The optimization program is likened to a problem to determine the dynamic eigenvalues of the spectral density matrix of the variables observed. The spectral matrix $I_x(\omega)$ of X_t is estimated using a representation of time series in the frequency domain for each frequency ω in the interval $[0, 2\pi[$. The estimated spectral matrix contains information on both the cross-correlation between variables and their dynamic relations. Thus, we write $\widehat{\Sigma}_x(\tau)$ the estimated autocovariance matrix between X_t and $X_{t-\tau}$ for a particular lag τ. The estimated spectral density of the vector of observed variables is given by:

$$\widehat{I}_x(\omega_h) = \sum_{\tau=-H}^{H} \widehat{\Sigma}_x(\tau) \left(1 - \frac{|\tau|}{H+1}\right) e^{-i\tau\omega_h} \tag{10.30}$$

for each Fourier frequency $\omega_h = 2\pi h/(2H+1)$ and for each $h = 0, ..., 2H$, with i representing the imaginary number such as $i^2 = -1$. For each frequency ω_h, the dynamic eigenvalues and eigenvectors resulting from $\widehat{I}_x(\omega_h)$ are calculated. The eigenvectors are classified in decreasing order. More specifically, the eigenvectors $\widehat{P}_l(\omega_h)$, of dimension $(N \times 1)$, are collected for $l = 1, .., q$ (the highest q eigenvalues). To return to the time domain, the eigenvectors are obtained based on the inverse Fourier transform:

$$\widehat{P}_l(L) = \sum_{\tau=-H}^{H} \widehat{P}_{l,\tau} L^{\tau}, \qquad \text{avec } \widehat{P}_{l,\tau} = \frac{1}{2H+1} \sum_{h=0}^{2H} \widehat{P}_l(\omega_h) e^{i\tau\omega_h} \tag{10.31}$$

for $\tau = -H, ..., H$ et $j = 1, ..., q$. The j-th dynamic principal component, $\widehat{F}_{j,t}$, is, then, given by the j-th component of $\sum_{l=1}^{q} \widehat{P}_l'(L)\widehat{P}_l(L)X_t$.

Thus, the dynamic principal components are obtained from a decomposition of the spectral density matrix into dynamic eigenvalues and eigenvectors. This breakdown also makes it possible to divide the spectral density matrix into a spectral density matrix of common components $I_\chi(\omega)$ and a spectral density matrix of idiosyncratic components $I_\xi(\omega)$.

Moreover, the estimator of the frequency domain is reduced to a symmetric filter. This presents problems at the end of samples, particularly when future observations are useful to estimate the principal components. To remedy this problem, Forni et al. (2005) suggest a refinement of their procedure that maintains the advantages of the dynamic approach, while basing the estimation of the common components on an asymmetric

[11] FHLR generalize the dynamic factor model of Sargent and Sims (1977) and Geweke (1977) by raising the hypothesis of the orthogonality of the idiosyncratic factors (see also Forni et al. (2009)). Hallin and Liska (2011) recently adapted these models to estimate common factors specific to "blocks" of data, i.e., large subpanels of variables.

filter.[12] With this procedure, the space of the factors is approximated by adding the r static factors rather than the q dynamic principal components. However, the mean resulting from the r contemporary static factors is based on information from the dynamic approach. The estimation of the model, then, consists in maximizing the variance of the common components or minimizing the variance of the idiosyncratic components. Thus, a convergent estimator of the spectral density matrix of the common component is given by:

$$\widehat{I}_{\chi}(\omega) = \widehat{P}(\omega)\Omega(\omega)\widehat{P}'(\omega), \tag{10.32}$$

where $\Omega(\omega)$ is a diagonal matrix of dimension $(q \times q)$ containing the largest q dynamic eigenvalues on the diagonal and $\widehat{P}(\omega) = (\widehat{P_1}(\omega), ..., \widehat{P_q}(\omega))$ is a matrix of dimension $(N \times q)$ containing the eigenvectors corresponding to the frequency ω. We, then, deduce the spectral density matrix of the idiosyncratic component:

$$\widehat{I}_{\xi}(\omega) = \widehat{I}_{x}(\omega) - \widehat{I}_{\chi}(\omega) \tag{10.33}$$

This estimation in the frequency domain is carried out in two steps. The first one is based on the dynamic approach, by which we obtain the variance-covariance matrices of the common components $\widehat{I}_{\chi}(\omega)$ and idiosyncratic components $\widehat{I}_{\xi}(\omega)$ respectively, estimated by an inverse Fourier transform. Thus, the variance-covariance matrix of the common components is estimated as follows:

$$\widehat{\Sigma}_{\chi}(\tau) = \frac{1}{2H+1}\sum_{h=0}^{2H}\widehat{I}_{\chi}(\omega_h)e^{i\tau\omega_h} \tag{10.34}$$

for $\tau = -H, \ldots, H$. The variance-covariance matrix of the idiosyncratic components is estimated in the same way:

$$\widehat{\Sigma}_{\xi}(\tau) = \frac{1}{2H+1}\sum_{h=0}^{2H}\widehat{I}_{\xi}(\omega_h)e^{i\tau\omega_h} \tag{10.35}$$

In a second step, this information is used to construct the factor space using the r aggregated means. More specifically, the variables are weighted in terms of the common-to-idiosyncratic variance ratio, obtained by means of the variance-covariance matrices estimated in the first step. These r aggregated means are defined as the solutions to a generalized principal components problem and they have the advantage of minimizing the idiosyncratic quadratic errors of the common factors by selecting only those variables with the highest common-to-idiosyncratic variance ratio. The number of these aggregated means is equal to $r = s(p+1)$, which represents the rank of the spectral density matrix of the common factors, where s indicates the number of lags for $\lambda_i^*(L)$ in equation (10.10). Forni et al. (2005) show that to determine the number of aggregated means r, the problem of maximization can be converted into a problem of generalized eigenvalues:

$$\hat{\Sigma}_{\chi}(0)\hat{Z}_j = \hat{\mu}_j\hat{\Sigma}_{\xi}(0)\hat{Z}_j \tag{10.36}$$

where $\hat{\mu}_j$ is the j-th generalized eigenvalue, $\hat{Z}_j$ its associated eigenvector of dimension $(N \times 1)$, and $\hat{\Sigma}_{\chi}(0)$ and $\hat{\Sigma}_{\xi}(0)$ are the contemporaneous variance-covariance matrices ($\tau = 0$) of the common and idiosyncratic components, respectively. Moreover, FHLR (2005) impose the following normalization $\hat{Z}_j'\hat{\Sigma}_{\xi}(0)\hat{Z}_{j'} = 1$ for $j = j'$ and $\hat{Z}_j'\hat{\Sigma}_{\xi}(0)\hat{Z}_{j'} = 0$ for $j \neq j$. Then, the eigenvalues are classified in decreasing order and the factors obtained correspond to the product of the r eigenvectors corresponding to the highest eigenvalues and the vector X_t. The estimator proposed by Forni et al. (2005) is written as follows:

$$F_t^{FHLR} = \widehat{Z}'X_t \tag{10.37}$$

where $\widehat{Z} = (\widehat{Z_1}, ..., \widehat{Z}_r)$ is a matrix of dimension $(N \times r)$ of the stacked eigenvectors.

[12] See also Forni and Lippi (2011).

To conclude this section, we will point out that the estimation methods proposed are relatively recent and the literature does not yet have sufficient perspective to systematically choose one method over another. Because it is easy to use, the Stock and Watson (2002) approach is naturally attractive, and the empirical results, particularly in a forecasting context, show that the results yielded by this approach are not significantly poorer than the other approaches in terms of forecasting error (on this point, see D'Agostino and Giannone (2012), or Barhoumi et al. (2013)).

The recent literature has also looked at the estimation of factor models in a Bayesian context. This approach makes it possible to reduce uncertainty regarding the parameters by first applying hypotheses on the distributions of these parameters. In this regard, we refer interested readers to Kose et al. (2003), Kose et al. (2008) or Lopes and West (2004), for examples.

Finally, the asymptotic properties of the estimators presented above are proved under the simple hypothesis « N and T tend to infinity » the interpretation of which is sometimes rather vague. Bai (2004) and Forni et al. (2004) emphasize that asymptotic properties, such as convergence, hold along specific trajectories $\{(N, T(N)); N \in \mathbf{N}\}$. For example, a property that holds for $\min(N, T)$ holds along the entire trajectory $(N, T(N))$, while a property that holds for $N = O(T^k)$ requires that the number of observations T be at least on the order of $N^{1/T}$. In fact, three concepts of limits exist: (i) sequential, (ii) pairwise, and (iii) simultaneous. Let $g(N, T)$ be a function that one wishes to study. A sequential limit stretches N and T to infinity, one after the other. A pairwise limit stretches (N, T) to infinity only along a particular trajectory, which can be denoted $\lim_{N,T\to\infty} g(N, T(N))$. A simultaneous limit authorizes (N, T) to increase along all possible trajectories: $\lim_{N,T\to\infty} g(N, T)$. It is noteworthy that the existence of a simultaneous limit implies the existence of a pairwise limit and a sequential limit, but the reverse is not true. Another approach using the theory of random matrices postulates that N and T tend to infinity with $N/T \to c \in (0, \infty)$, where c is a constant. For a more detailed discussion, see, for example, Bai and Ng (2008b) and Harding (2009).

10.5 Selection of the number of factors

An important step in the statistical analysis of static and dynamic factor models is the preliminary identification of the number of factors. A number of papers focus on the problem of determining the number of factors. For example, Forni and Reichlin (1998) suggest a graphic approach to identify the number of factors when $N \to \infty$ and T is fixed but no theory is proposed. Stock and Watson (1998) modify the BIC criterion to select the optimal number of factors in forecasting when $N, T \to \infty$ with $\sqrt{N}/T \to \infty$.However, their criterion is restrictive since it requires that $N \gg T$, and it is appropriate only in a forecasting context. Forni et al. (2000) consider a multivariate version of the AIC criterion but no theoretic or empirical property is known for their criterion.

In this section, we present the criteria most used in the empirical literature, i.e., the criteria of Bai and Ng (2002) and Alessi et al. (2010) for static factor models and those of Stock and Watson (2005a), Amengual and Watson (2007), Bai and Ng (2007), Hallin and Liska (2007) and Breitung and Pigorsch (2013) for dynamic factor models. Note that those criteria have been compared in a forecasting framework by Barhoumi et al. (2013).

10.5.1 Selection of the number of factors for static factor models

To specify the number of factors, Bai and Ng (2002) suggest using information criteria to select the optimal number of static factors r, when N and T tend to infinity. Bai and Ng (2002) propose information criteria based on the quality of adjustment of the model to the data measured by the variance $V(j, F)$ such that :

$$V(j, F) = (NT)^{-1} \sum_{t=1}^{T} \left(X_t - \widehat{\Lambda}\widehat{F}_t \right)^2, \tag{10.38}$$

where j is a given number of factors such as $\widehat{F}_t = (\widehat{F}_{1t}, \ldots, \widehat{F}_{jt})'$. Thus, if the number of factors j increases, the variance of the factors increases mechanically and the sum of the squares of the residuals decreases in turn. Bai and Ng (2002), then, suggest introducing a penalty function in the criterion to be optimized and propose the following three criteria, corresponding to different penalty functions:

$$IC_1(j) = \ln\left(V(j,F)\right) + j.\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right), \tag{10.39}$$

$$IC_2(j) = \ln\left(V(j,F)\right) + j.\left(\frac{N+T}{NT}\right), \tag{10.40}$$

$$IC_3(j) = \ln\left(V(j,F)\right) + j.(\ln C_{NT}^2/C_{NT}^2), \tag{10.41}$$

where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$ and $\ln$ denotes the natural logarithm. The estimation of the number of factors r is obtained by minimization of the information criteria for $j = 0$, $j = r_{\max}$, where $r_{\max}$ is the maximum number of static factors. These criteria reflect the trade-off between the quality of the adjustment and the risk of overadjustment[13]. Bai and Ng (2002) show that their criteria are robust to the presence of a heteroskedastic component in the time and cross-section dimensions between variables, but also in the presence of weak serial and cross-section dependence.

Subsequently, Alessi et al. (2010) extend this criterion by modifying the strength of the penalty function that appears in the preceding three criteria given by equations (10.39)-(10.40)-(10.41). Alessi et al. (2010) propose an alternative to the criteria proposed by Bai and Ng (2002) by multiplying the penalty function by a positive constant c suggested originally by Hallin and Liska (2007), representing the strength of the penalty function. The authors, thus, propose the following two criteria:

$$IC_1^*(j) = \ln\left(V(j,F)\right) + c.j.\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right) \tag{10.42}$$

$$IC_2^*(j) = \ln\left(V(j,F)\right) + c.j.\left(\frac{N+T}{NT},\right) \tag{10.43}$$

where $V(j,F)$ is given by equation (10.38). The estimation of the number of factors r is obtained by minimization of the information criteria IC_1^* and IC_2^* for $j = 0$, $j = r_{\max}$, where $r_{\max}$ is the maximum number of static factors.

The procedure for the selection of the number of static factors depends both on the variance of the number of estimated factors $V_c(r)$ (for N and T tending to infinity) and on the constant $c \in [0, c_{\max}]$. Alessi et al. (2010) suggest estimating this variance $V_c(r)$ by reiterating the procedure for estimating r for a finite number of subsets of the initial N variables, also making the number of observations T vary.

Kapetanios (2010) proposes a concurrent method to the information criterion to estimate the number of static factors, based on the random matrix theory. His approach is based on a series of tests on the largest eigenvalues of the variance-covariance matrix of the initial data, which we have denoted Σ_X. Other procedures have been suggested by Yao and Pan (2008) and Onatski (2010).

[13] Bai and Ng (2002) also propose another class of information criteria for which the variance $V(j,F)$ replaces $\ln\left(V(j,F)\right)$ in equations (10.39), (10.40) and (10.41). Bai and Ng (2002), Theorem 2, give the results of the convergence of these criteria when N and T tend to infinity.

10.5.2 Selection of the number of factors for dynamic factor models

The Bai and Ng (2007) criterion

In the context of dynamic factor models, the number of dynamic shocks q (for the estimation of factors in dynamic principal components and their space-state form) can be determined using the Bai and Ng (2007) information criterion. This criterion is obtained by considering the r estimated static factors as given and, then, estimating a VAR model of order p on these factors, where the order p is selected using the BIC criterion. Next, a spectral decomposition of the variance-covariance matrix of the estimated residuals of the VAR model, denoted $\widehat{\Sigma}_\varepsilon$ of dimension $(r \times r)$, is calculated. Then, the j-th ordered eigenvalue $\widehat{c}_j$, where $\widehat{c}_1 > \widehat{c}_2 \geq ...\widehat{c}_r \geq 0$ is recovered. Finally, for $l = 1, ..., r-1$, Bai and Ng (2007) propose the following two quantities:

$$\begin{aligned} \widehat{D}_{1,l} &= \left(\frac{\widehat{c}_{l+1}}{\sum_{j=1}^{r} \widehat{c}_j} \right)^{1/2} \\ \widehat{D}_{2,l} &= \left(\frac{\sum_{j=l+1}^{r} \widehat{c}_j}{\sum_{j=1}^{r} \widehat{c}_j} \right)^{1/2} \end{aligned}$$

where $\widehat{D}_{1,l}$ represents a measure of the marginal contribution of the $l+1^{me}$ eigenvalue and $\widehat{D}_{2,l}$ represents a measure of the cumulative contribution of the eigenvalues, under the hypotheses that $\widehat{\Sigma}_\varepsilon$ is the unit matrix of dimension $(r \times r)$ and that $c_l = 0$ for $l > q$[14].

Thus, according to the selected marginal contribution measure, the number of dynamic factors q is obtained by minimizing:

$$\left\{ l \quad \text{tel que} : \widehat{D}_{1,l} \leq \frac{c}{\min\left[n^{\frac{2}{5}}, T^{\frac{2}{5}}\right]} \right\},$$

or:

$$\left\{ l \quad \text{tel que} : \widehat{D}_{2,l} \leq \frac{c}{\min\left[n^{\frac{2}{5}}, T^{\frac{2}{5}}\right]} \right\}.$$

Bai and Ng (2007) suggest using $c = 1$ based on Monte Carlo simulations.

In practice, these different criteria are used at three stages:

1. First, one of the Bai and Ng (2002) criteria is used to determine the optimal number of factors $r \in \{1, \ldots, r_{max}\}$ in a static context[15] ;
2. Then, a VAR(p) is estimated on these r estimated factors and the order p of the VAR is selected to minimize the BIC criterion;
3. Finally, the Bai and Ng (2007) criteria are applied to the variance-covariance matrix or correlation matrix of the residuals (ε_t) of the VAR(p) to obtain the optimal number of dynamic factors q.

The Stock and Watson (2005a) and Amengual and Watson (2007) criteria

Stock and Watson (2005a) and Amengual and Watson (2007) show that the Bai and Ng (2002) estimator can be used to estimate the number of dynamic factors. To do this, they propose applying this estimator to the errors resulting from the projection of observed data on the lagged values of static factors, i.e., on

[14] Bai and Ng (2007) show that $\widehat{D}_{1,l}$ and $\widehat{D}_{1,l}$ converge toward zero when $l > q$.
[15] The criterion IC_2 is used more often in practice.

$\hat{\nu}_t = X_t - \sum_{\tau=1}^{p} \Lambda\Phi(L)F_{t-\tau}$. They propose two ways of estimating the errors (ν_t):

$$\begin{aligned}\widehat{\nu}_t^A &= X_t - \sum_{\tau=1}^{p} \widehat{\Lambda}\widehat{\phi}_\tau \widehat{F}_{t-\tau} \\ \widehat{\nu}_t^B &= X_t - \sum_{\tau=1}^{p} \widehat{\Pi}_\tau \widehat{F}_{t-\tau}\end{aligned}$$

where $(\widehat{\phi}_1, \widehat{\phi}_2, \ldots, \widehat{\phi}_p)$ are the ordinary least squares estimators of the regression of $\widehat{F}_t$ on $(\widehat{F}_{t-1}, \ldots, \widehat{F}_{t-p})$ and $(\widehat{\Pi}_1, \widehat{\Pi}_2, \ldots, \widehat{\Pi}_p)$ are the ordinary least squares estimators of the regression of X_t on $(\widehat{F}_{t-1}, \ldots, \widehat{F}_{t-p})$.

The Breitung and Pigorsch (2013) criteria

Breitung and Pigorsch (2013) also propose two information criteria to select the number of dynamic factors. Their criteria are based on an analysis of the canonical correlations of static factors (obtained by a principal components analysis) and depend on the estimation of a VAR(p) model on these factors, where the order p is selected by the BIC criterion.

The first criterion is based on the following statistic:

$$\zeta(q^*) = \tilde{C}_{NT}^{2-\delta} \sum_{j=1}^{r-q^*} (1 - \tilde{\lambda}_j)$$

where $\tilde{C}_{NT}^{2-\delta} = (2-\delta)N^{-1} + (2-\delta)T^{-1}$, with $0 < \delta < 2$, and $\tilde{\lambda}_j$ are values resulting from the solution to the following problem $|\tilde{\lambda}_j \tilde{S}_{00} - \tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{01}'| = 0$, with $\tilde{S}_{00} = \sum_{t=\tau+1}^{T} \hat{F}_t \hat{F}_t'$, $\tilde{S}_{01} = \sum_{t=\tau+1}^{T} \hat{F}_t \hat{G}_{t-1}'$, $\tilde{S}_{11} = \sum_{t=\tau+1}^{T} \hat{G}_{t-1}\hat{G}_{t-1}'$, and $\hat{G}_{t-1} = \left[\hat{F}_{t-1}', \ldots, \hat{F}_{t-\tau}'\right]$. The number of dynamic factors can be estimated using a large number of q^* resulting from this sequence $q^* = r-1, r-2, \ldots, 0$, where the statistic $\zeta(q^*)$ is larger than the level of the threshold κ, or:

$$q = \max\{q^* \quad \text{such that :} \quad \zeta(q^*) > \kappa\}$$

Breitung and Pigorsch (2013) suggest using the following values of the parameters: $\tau = 1$, $\delta = 0,5$ and $\kappa = 1$.

The second criterion is based on the following statistic:

$$LR(q^*) = T \sum_{j=r-q^*+1}^{r} \tilde{\lambda}_j$$

where the null hypothesis is H_0: $q^* = q$ as against the alternative H_1: $q^* < q$.

The Hallin and Liska (2007) criterion

Hallin and Liska (2007) develop an information criterion for generalized dynamic factor (GDF) models. This criterion is based on the spectral density matrix of the observations. It is written as follows:

$$IC_{2,N}^T(j) = \ln\left[\frac{1}{N}\sum_{s=j+1}^{N} \frac{1}{2M_T+1} \sum_{h=-M_T}^{M_T} \lambda_{N_s}^T(\omega_h)\right] + c.j.p(N,T), \tag{10.44}$$

with $0 \leq j \leq q_{\max}$, $\omega_h = \pi h/(M_T + \frac{1}{2})$ for $h = -M_T, \ldots, M_T$ with the truncation parameter $M_T > 0$, c is a positive constant such that $c = [0,01; 0,02; \ldots; 3,00]$, $N_s < N$ is the number of variables contained in a given subset and $p(N,T)$ is a penaly function[16] such that:

$$p(N,T) = \left(-M_T^{-2} + M_T^{1/2} T^{-1/2} + N^{-1} \right) \times \ln \left(\min[N, M_T^2, M_T^{-1/2} T^{1/2}] \right)$$

The eigenvalues $\lambda_{N_s}^T$ result from $\hat{I}_x(\omega)$, which represents the spectral density matrix estimator of X_t with $\omega \in [-\pi, \pi]$. The number of estimated factors is, then, given by:

$$q = \underset{0 \leq j \leq q_{\max}}{\text{argmin}}\, IC_{2,n}^T(j)$$

The procedure for selecting the number of dynamic factors is similar to that used by Alessi et al. (2010), i.e., by examining the variance of the number of estimated factors, $V_c(r)$ for n and T tending to infinity and for an interval of values for the constant c. In their numerical illustration, Hallin and Liska (2007) propose retaining $M_T = [0,75\sqrt{T}]$ and $q_{\max} = 13$.

10.6 Recent results in the empirical literature

Applications of dynamic factor models abound in the empirical economic literature. A few examples are asset pricing models (Ross (1976), consumer theory (Gorman (1981); Lewbel (1991)), and the assessment of performance and risk measurement in finance (Campbell et al. (1997)). In this section, we present some recent applications of these models that underscore the interest of this approach for (i) the construction of short-term economic indicators, (ii) macroeconomic forecasting, and (iii) international macroeconomics and monetary policy analysis.

10.6.1 Tools for short-term economic monitoring

Dynamic factor models are useful for developing economic activity indicators based on the mass of data available to short-term forecasters. These models can be used to synthesize large datasets into a composite indicator that reflects the most relevant data available on a given date.

One of the most common indicators to which forecasters working on the U.S. economy refer is the Chicago Fed National Activity Indicator (CF-NAI) developed using the Stock and Watson (1999) approach. This indicator is based on 85 monthly series representative of the U.S. economy, covering production, income, employment, personal consumption, housing, sales, inventories and orders. The CF-NAI corresponds to the first factor estimated in a principal components analysis. A value close to zero means that activity is close to its long-term trend. As well, the CF-NAI can be used to detect recessions in the U.S. by using historical estimated thresholds. The Philadelphia Federal Reserve, for its part, each week publishes a daily economic activity indicator that is based on the mixed frequency model presented in the article by Aruoba et al. (2009) and includes daily, weekly, monthly and quarterly data. Given its high frequency, this indicator is interesting for forecasting since it makes it possible to provide an early signal very rapidly.

In the euro zone, the Centre for Economic Policy Research (CEPR) has for some years now been disseminating the EuroCoin indicator developed at the Bank of Italy by Altissimo et al. (2001) and Altissimo et al. (2010). The purpose of this indicator is to estimate a monthly GDP growth in the euro zone for the coincident quarter using smoothing to remove very short-term effects (frequency less than one year). The model used to synthesize the information is a generalized dynamic factor model proposed by Forni et al. (2000) applied

[16] Hallin and Liska (2007) also propose two other penalty functions. Onatski (2009) suggests alternative tests in the context of approximate dynamic factor models. Jacobs and Otter (2008) propose a test based on a canonical correlation procedure to determine simultaneously the number of dynamic factors q and the order of the lag p in a dynamic factor model, but for fixed N and T.

to a very large number of variables. From an economic standpoint, this measure corresponds to a kind of medium-term quarterly growth, but does not aim to precisely estimate the quarterly accounts figures provided by Eurostat. The first version of EuroCoin has a smaller variance than quarterly GDP growth. The economic outlook that it provides deviates considerably from the national accounts. CEPR has recently tried to construct a more effective new version of EuroCoin (Alessi et al. (2010)). The aim of the new indicator is the same but the number of input variables used has been reduced from 951 to 145 (IPI, monetary aggregates, interest rates, financial variables, demand indicators, surveys, trade variables and labor market variables). The input variables have not been smoothed by statistical means, which eliminates some side effects that might have introduced a bias into the old EuroCoin.

For France, Doz and Lenglart (1999) have developed a summary indicator for the industry tendency survey conducted by INSEE by means of six monthly balances of opinion. This approach is now commonly used by INSEE to calculate composite indicators based on data from short-term tendency surveys in the various sectors. Clavel and Minodier (2009) extend this business activity indicator relating to industry by proposing a business climate indicator for the economy as a whole, which incorporates balances of opinion derived from the INSEE tendency surveys conducted in the various sectors such as services, construction, and wholesale and retail sales.

Many economic activity indicators have been developed on the basis of regime-switching factor models. Kim and Nelson (1998) propose an application for the four main U.S. economic series (growth rates of IPI, employment, income, and retail sales), also considered by the National Bureau of Economic Research (NBER) for the dating of U.S. recessions. These same series are considered by Diebold and Rudebusch (1996), Chauvet (1998) and Chauvet and Piger (2008), who simultaneously estimate a similar model to construct a real-time business cycle indicator. In France, Nguiffo-Boyom (2006) estimates the Kim and Nelson (1998) model simultaneously on four series derived from the INSEE tendency surveys to reproduce the growth cycle, which measures the long-term deviation from trend. For the euro zone, Darné and Ferrara (2011) propose a two-step dynamic factor model. They, first, estimate a factor which, in turn, follows a regime-switching model, based on a set of six business confidence index series for the six main eurozone countries. The authors, thus, develop an indicator to detect the acceleration cycle in the monetary zone in real time.

As for the mixed frequency approach, Mariano and Murasawa (2003) have applied their model to estimate an economic activity indicator in the U.S. Mariano and Murasawa (2010) also use a version of their model to calculate a monthly U.S. GDP series by estimating a mixed frequency VAR that includes the quarterly GDP growth series and the four monthly series traditionally used by NBER to assess the U.S. business cycle (growth rates of employment, income, industrial production, and retail sales). Cornec (2006) also uses this approach, first to provide a monthly dating of the French business cycle using two quarterly series (GDP growth rate and employment) and two monthly series (IPI and household consumption expenditure), and then to estimate a composite activity indicator similar to that of Doz and Lenglart (1999) but which includes the quarterly GDP growth rate series as supplementary information. The empirical results of this application emphasize that the contribution of GDP to the first factor is negligible in comparison with the composite business indicator relating to industry. Again in France, Cornec and Deperraz (2006) use a mixed frequency model to develop an activity indicator in the service sector based on three monthly balances and three quarterly balances from the INSEE tendency survey relating to services. This indicator can usefully supplement the indicator on the business climate in manufacturing for short-term forecasting. Clavel and Minodier (2009) also develop a mixed frequency approach to incorporate the various short-term tendency surveys conducted by INSEE, which are sampled monthly, bimonthly and quarterly, into their business climate indicator. More recently, Camacho and Perez-Quiros (2010) and Camacho and Perez-Quiros (2011) develop two short-term growth indicators for Spain and the eurozone using a mixed frequency factor model that also includes Markov regime-switching to take account of the business cycle. Frale et al. (2010) develop a small-dimension mixed frequency factor model to provide a measure of monthly GDP in the eurozone. Finally, Frale et al. (2011) propose a mixed frequency model to estimate a monthly GDP indicator in the eurozone called EUROMIND, based on a disaggregation between supply and demand. This indicator is based on the official database for the eurozone

developed by Eurostat (Euro-IND).

10.6.2 Macroeconomic forecasting

Dynamic factor models are widely used, particularly by central banks, as a tool for forecasting various macroeconomic variables, such as the GDP growth rate or inflation (see, for example, a survey in Stock and Watson (2006), or Eickmeier and Ziegler (2008). When the forecasting horizon covers the current period, such forecasting is termed "nowcasting" (on this point see Giannone et al. (2008)). The factors are estimated from monthly data used to track countries' economic situation, such as household and business survey data (soft data), variables for the real economy (hard data), including indices of industrial production, household consumption, retail sales or new vehicle registrations and, finally, financial variables (stock prices, oil prices, interest rates, etc.). For a given country, such a database can include several hundred variables. It is useful, therefore, to be able to synthesize this large data set into a small dimension vector to be included in standard models.

Based on asymptotic theoretical results on the convergence of estimators in this type of model, early work used the largest possible number of variables available. More recent work addresses the question whether including the largest number of variables is appropriate or not to improve the accuracy of forecasts. For example, Barhoumi et al. (2010) show empirically in the case of France that increasing the data set by disaggregation does not result in significant improvements in the accuracy of short-term GDP forecasts. Boivin and Ng (2006) identify the conditions under which expanding the database could result in less accurate factor estimates and provide empirical rules for eliminating redundant variables. These authors show that expanding the data set is not preferable if the new series add too much idiosyncratic noise and/or increase the cross correlation between idiosyncratic errors too much. Bai and Ng (2008) use LARS-type statistical methods (least angle regressions), which are weighted regressions, to identify optimal subsets of predictors (targeted predictors) from a large dataset. Schumacher (2010) underscores the effectiveness of this approach in using an international database to predict German growth, which is very sensitive to fluctuations in the international environment. Charpin (2009) application of this approach to French data also seems to provide encouraging results.

Once the factors have been estimated, the forecasting of the variable of interest Y_t over a horizon h is derived from either an ARDL-type univariate regression equation (autoregressive distributed lags, see equation (10.45) below) or a VAR-type multivariate process. When the aim is to forecast over a horizon h exceeding one step, two approaches co-exist: the recursive approach, which uses, for a given step, the forecasts made for the previous steps, and the direct approach, which seeks to predict the value over the horizon h directly, i.e. without trying to forecast the variable of interest in the previous steps. In a general context, direct forecasting of the variable over the horizon $h$makes it possible to reduce the forecast bias resulting from the estimation of the parameters that may appear in the case of a multistep recursive forecast (see, for example, Chevillon (2007)). In the particular context of factor models, according to simulations done by Boivin and Ng (2005), there does not seem to be a significant difference between a direct forecast and a recursive forecast if estimated factors are used. However, the direct approach is preferred in many applications.

Thus, the univariate equation of direct forecasting to a horizon h is written:

$$\hat{Y}_{t+h|t} = \hat{\alpha}_h + \sum_{j=1}^{m} \hat{\beta}'_{hj} \hat{F}_{t-j+1} + \sum_{j=1}^{p} \hat{\phi}_{hj} Y_{t-j+1}, \tag{10.45}$$

Where $\hat{F}_t$ is the vector of dimension r of the estimated factors, m and p are the autoregressive orders, and $\hat{\beta}_{hj}$ is a vector of estimated coefficients of dimension r. The parameters α_h, β_{hj} and ϕ_{hj} depend on the horizon h since, in the context of a direct forecast, they vary on the basis of the horizon considered. The $mr + p + 1$ parameters of the model are estimated using ordinary least squares. In equation (10.45), the number of factors r can be specified by one of the tests presented above. However, $r = 3$ is often used in

practice, since three factors are often sufficient to explain a significant portion of the data variance. Three variants of the model given in equation (10.45) are generally used (see Stock and Watson (2002), Boivin and Ng (2005). The first, denoted DI (*Diffusion Index*) is obtained with $m = 1$ and the terms depending on $p = 1$ being suppressed in (10.45) and, thus, includes only contemporaneous information $\hat{F}_t$. The second, *DI-AR*, authorizes a dynamics on the series Y_t and corresponds to $m = 1$ and $1 \leq p \leq 6$ in (10.45). The optimal autoregressive order p is, then, obtained by minimization of an AIC or BIC-type information criterion. Finally, the specification *DI-AR,Lag* of (10.45) corresponds to $1 \leq m \leq 3$ and $0 \leq p \leq 3$, thus allowing lags on the factors and on the variable Y_t.Once again, the optimal parameters m and p are obtained by minimization of an information criterion. We should point out that the specification *DI-AR,Lag* is not used for a dynamic factor since such a factor is assumed to already include a time dynamics.

Boivin and Ng (2005) use a simulation to show that the differences between using static or dynamic factors are negligible in forecasting. Barhoumi et al. (2010) find the same result empirically on French data. Barhoumi et al. (2010) also show that the specification of the model used to make the forecasts has only a marginal impact on the quality of the forecast, particularly when the number of observations is high.

One of the major problems that appears when these models are used for real-time forecasting results from the fact that data arrive in a staggered fashion, leading to missing values at the end of samples (this is the ragged-edge data problem, which is well-known from forecasters). Several solutions have been proposed in the empirical literature, such as the projection of missing data, either using an autoregressive-type parametric model or using moving averages, or the realignment of the data base on the last points available, if the number of variables is high. In the context of factor models, two-step estimation using a Kalman filter (Doz et al. (2011)) solves this problem elegantly (Giannone et al. (2008), or Angelini et al. (2011)).

Among the many applications in forecasting the GDP growth rate, we can cite, for example, the articles of Stock and Watson (2002) or Banerjee and Marcellino (2006) for the U.S., those of Barhoumi et al. (2010), Barhoumi et al. (2013) and Bessec and Doz (2012) for France, Forni et al. (2000), Forni et al. (2003), Camba-Méndez et al. (2005), Marcellino et al. (2005), Banerjee et al. (2005), Ruenstler et al. (2009) or Angelini et al. (2011) for the eurozone, Schumacher (2007), Schumacher (2010), Schumacher and Breitung (2008) and Eickmeier and Ziegler (2008) or Marcellino and Schumacher (2010) for Germany, Artis et al. (2005) for the United Kingdom, and Van Nieuwenhuyze (2006) for Belgium. Matheson (2014) also develops GDP forecasts for a large number of advanced and emerging countries.

It should be noted that applications to inflation forecasting are much rarer; see, for example, Forni et al. (2003) or Camba-Méndez et al. (2005) for the eurozone or de Bandt et al. (2007) for France. Boivin and Ng (2005) also consider series of U.S. prices from a forecasting point of view. It appears to be difficult to improve the accuracy of inflation forecasts by using a large number of variables in comparison with an approach based on a precise selection of variables of interest. In contrast, measurements of underlying inflation have been conducted using this type of approach; we refer interested readers to the articles of Cristadoro et al. (2005) for the eurozone or Kapetanios (2004) for the United Kingdom.

Various works have endeavored to identify the contribution of financial variables to macroeconomic forecasting using factor models applied to a database on financial markets activity. For example, Forni et al. (2003) show that, in the eurozone, financial variables help to forecast inflation but cannot be used to accurately forecast industrial output. Bellégo and Ferrara (2009) and Bellégo and Ferrara (2012) use a factor model to assess the likelihood of a recession in the eurozone based on a large set of monthly variables (factor-probit-type model)[17] Specifically, Bellégo and Ferrara (2009) show that, by using only financial variables, this approach would have made it possible to anticipate the 2008-09 recession in the eurozone in real time as early as late 2007.

A review of the literature on the results of factor models in forecasting is found in the article by Eickmeier and Ziegler (2008), who conduct a meta-analysis of the performance of models to forecast GDP and inflation.

[17] A factor-probit model is obtained by, first, estimating the factors from a data set, and then, incorporating them in a standard probit-type model.

They conclude that factor models generally improve smaller-scale econometric models, but that methods of combining forecasts[18] constitute a competing alternative approach.

10.6.3 Applications in monetary policy and the international economy

An abundant literature analyzes the impact of monetary policy shocks on the macroeconomy and how the transmission mechanism of these shocks has evolved, particularly for the U.S. Traditionally, the impact of monetary policy shocks is often measured using small dimension VAR models (fewer than 6 variables). Typically, trivariate SVAR models containing the interest rate, output and inflation are used. Beyond this small number of variables, it is difficult to estimate this type of model using standard methods (in a Bayesian context, see, for example, De Mol et al. (2008)). Now, as noted in Bernanke and Boivin (2003), monetary policy is conducted by central banks in a data-rich environment. In this context, a category of papers analyzing monetary policy initiated by the article of Bernanke et al. (2005) uses FAVAR models. This type of modeling is chosen to remedy the problem of missing variables generally encountered in traditional VAR modeling. Bernanke et al. (2005), Stock and Watson (2005a) and Favero et al. (2005) use FAVAR models to analyze monetary policy in the U.S. and in some eurozone countries. They all conclude that adding estimated factors from factor models to VAR models allows for a finer-grained analysis of the phenomena in question, particularly in terms of structural shocks. For example, Del Negro et al. (2007) estimate a factor common to changes in residential housing prices in various U.S. states, then introduce it into a FAVAR to assess the extent to which monetary easing helped create a housing bubble (the study data stop in 2005). They show that the impact of monetary policy shocks is weak in comparison with the scope of the price fluctuations observed through to the end of their sample.

Another portion of the literature addresses the question of whether the mechanism for the transmission of shocks has changed over time and, if so, how. Time-varying FAVAR (TV-FAVAR) models provide an extremely flexible modeling that can shed light on this question. The literature appears to be in agreement on the fact that this mechanism has changed, although there is no consensus on how. For example, based on a set of 803 quarterly variables from 1972 to 2007, Eickmeier et al. (2015) show that the volatility of monetary shocks in the U.S. declined significantly from the early 1980s to the eve of the subprime crisis and that the negative impact of a shock on U.S. activity and prices declined over this period. The authors also emphasize that the negative impact of a monetary policy shock on inflation expectations and long-term interest rates weakened over time. The reasons given by the authors for these evolutions are the changes in monetary policy and the globalization of trade and finance. Finally, these authors indicate that the transmission mechanism appear to be the same in periods of expansion and periods of contraction. Baumeister et al. (2010) also show that, for the U.S. economy, the reaction of GDP, consumption and investment to a monetary shock declined over the period 1960-2008. We should point out, however, that most of these studies do not include the period 2008-2009, during which the industrialized countries suffered the worst recession since the 1920s.

FAVAR models have also been used to analyze changes in the synchronization of global business cycles, making it possible to discriminate between different types of shocks. For example, Stock and Watson (2005b) estimate a FAVAR on the GDP of the G7 countries, enabling them to identify common international shocks and domestic effects due to an international shock and those due to an idiosyncratic shock. They conclude that the reduction in the volatility of cycles in the G7, with the exception of Japan, observed between the mid-1980s and the mid-2000s, is primarily due to a reduction in the amplitude of common international shocks (the so-called “Great Moderation”). Kose et al. (2003) consider a similar model to show the existence of a global cycle based on a set of 60 countries. They also show that factors specific to the region play only a minor role in the explanation of macroeconomic fluctuations. Bordo and Helbling, (2010) explore the historical angle by using annual GDP data from 1880 to 2008 for 16 industrialized countries and show a trend toward increasing synchronization among those countries. The authors show the role of common shocks in this change using a restricted FAVAR model estimated on this database.

[18] Forecast combinations generally use a weighted average of forecasts of a single target variable, based on a large number of different models. See Timmerman (2006).

There are also several applications of factor models to the measurement of international cycles and their transmission between countries. Mansour (2003) and Helbling and Bayoumi (2003) estimate a global business cycle for the world and the G7 countries and analyze the contribution of this common cycle to economic changes in each country. Kose et al. (2008) use a factor model in a Bayesian context to estimate the common and idiosyncratic components of the G7 countries for a set of economic aggregates. They show that the factor common to these countries explains a larger portion of the variation of these aggregates over the period 1986-2003 than in previous periods, thus demonstrating an increased synchronization of the business cycles in the G7. Eickmeier, (2007) analyzes the transmission of U.S. structural shocks to Germany using the approach put forward by Forni et al. (2004). After analyzing common economic movements in the eurozone, Marcellino et al. (2005) and Eickmeier (2005) try to give economic interpretations to the common factors by linking them to the various countries in the zone and/or certain variables, based on correlation measures.

10.7 Conclusion

In this article, we have reviewed the recent literature on dynamic factor models. There has been recently an increasing interest in these models on the part of researchers since they can adequately respond to certain problems encountered in practice, particularly inflation of the number of available data. We have presented the models and their most interesting extensions, main estimation methods, and tests of the number of factors. In the final section, we have presented a few recent examples of the application of dynamic factor models to macroeconomic forecasting, the construction of short-term economic indicators and the analysis of monetary policy and the international economy. The success of dynamic factor models means that this review of the literature cannot claim to be exhaustive. Extensions have been developed very recently, particularly to add flexibility by means of non-linearities or mixed frequencies, and numerous applications continue to be published. Moreover, these models are now increasingly weighed against other econometric methods that also make it possible to reduce the scale of the problem. It appears to us, therefore, that the research into dynamic factor models will continue to thrive.

Bibliography

Alessi L., Barigozzi M. and Capasso M. (2010). Improved penalization for determining the number of factors in approximate factor models. Statistics and Probability Letters, 80, 1806-1813.

Altissimo F., Bassanetti A., Cristadoro R., Forni M., Hallin M., Lippi M. and Reichlin L. (2001). EuroCOIN: A real time coincident indicator of the euro area business cycle. Discussion Paper No. 3108, CE PR.

Altissimo F., Cristadoro R., Forni M., Lippi M. and Veronese G. (2010). New EuroCOIN: Tracking economic growth in real time. The Review of Economics and Statistics, 92, 4, 1024-1034.

Amengual D. and Watson M.W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. Journal of Business and Economic Statistics, 25, 91-96.

Anderson T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Angelini E., Camba-Méndez G., Giannone D., Rünstler G. and Reichlin L. (2011). Short-term forecasts of Euro area GDP growth. Econometrics Journal, 14, 1, C25-C44.

Artis M., Banerjee A. and Marcellino M. (2005). Factor forecasts for the UK. Journal of Forecasting, 24, 279-298.

Aruoba S., Diebold F.X. and Scotti C. (2009). Real-time measurement of business conditions. Journal of Business and Economic Statistics, 27, 4, 417-27.

Bai J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. Journal of Econometrics, 122, 137-183.

Bai J. and Ng S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70, 191–221.

Bai J. and Ng S. (2007). Determining the number of primitive shocks in factor models. Journal of Business and Economic Statistics, 25, 52-60.

Bai J. and Ng S. (2008). Forecasting economic time series using targeted predictors. Journal of Econometrics, 146, 304-317.

Bai J. and Ng S. (2008b).

de Bandt, O., E. Michaux, C. Bruneau and A. Flageollet (2007). Forecasting inflation using economic indicators: The case of France, Journal of Forecasting, 26, 1, 1-22.

Banerjee A. and Marcellino M. (2006). Are there any reliable leading indicators for US inflation and GDP growth? International Journal of Forecasting, 22, 137-151.

Banerjee A. and Marcellino M. (2009). Factor-augmented error correction models. In Castle, J. and Shepard, N. (Eds.), *The Methodology and Practice of Econometrics - A Festschrift for David Hendry* , Oxford: Oxford University Press.

Banerjee A., Marcellino M. and Masten I. (2005). Leading indicators for Euro area inflation and GDP growth. Oxford Bulletin of Economics and Statistics, 67, 785-813.

Barhoumi K., Darné, O. and Ferrara, L. (2010). Are disaggregate data useful for forecasting French GDP with dynamic factor models? Journal of Forecasting, 29, 1-2, 132-144.

Barhoumi K., Darné, O. and Ferrara, L. (2013). Testing the number of factors for dynamic factor modelling: An empirical assessment for forecasting purpose. Oxford Bulletin of Economics and Statistics, 75, 1, 64-79.

Baumeister C., Liu, P. and Mumtaz, H. (2010), Changes in the transmission of monetary policy: Evidence from a time-varying factor-augmented VAR. Working Paper No. 401, Bank of England.

Bellégo C. and Ferrara, L. (2009). Forecasting euro area recessions using time-varying binary response models for financial variables. Working Paper No. 259, Banque de France.

Bellégo C. and Ferrara, L. (2012). A factor-augmented probit model for business cycle analysis. Economic Modelling, 29, 1793-1797.

Bernanke B.S. and Boivin J. (2003). Monetary policy in a data-rich environment. Journal of Monetary Economics, 50, 525.

Bernanke B. S., Boivin J. and Eliasz P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. Quarterly Journal of Economics, 120, 387-422.

Bessec M. and Doz C. (2012). Prévision de court terme de la croissance du PIB français à l'aide de modèles à facteurs dynamiques. Économie et Prévision, 199, 1-30.

Boivin J. and Ng S. (2005). Understanding and comparing factor-based forecasts. International Journal of Central Banking, 1, 117-151.

Boivin J. and Ng S. (2006). Are more data always better for factor analysis? Journal of Econometrics, 132, 169-194.

Boivin J., Giannoni M.P. and Mihov I. (2009). Sticky prices and monetary policy: Evidence from disaggregated U.S. data. American Economic Review, 99, 350-384.

Bordo M. and Helbling, T. (2010). International business cycle synchronization in historical perspective. Working Paper 16103, NBER.

Breitung J. and Pigorsch U. (2013). A canonical correlation approach for selecting the number of dynamic factors. Oxford Bulletin of Economics and Statistics, 75, 23-36.

Brillinger D.R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco.

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

Camacho M. and Perez-Quiros G. (2011). Spain-STING: Spain short-term indicator of growth. The Manchester School, 79, 594-616.

Camba-Méndez G. and Kapetanios G. (2005). Forecasting euro area inflation using dynamic factor measures of underlying inflation. Journal of Forecasting, 25, 491-503.

Campbell J.Y., Lo A.W. and MacKinlay A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Chamberlain G. and Rothschild (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica, 51, 1305–1324.

Charpin, F. (2009). Estimation précoce de la croissance. De la régression LARS au modèle à facteurs. Revue de l'OFCE, 2009/1, 108, 31-48.

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

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.

Chevillon G. (2007). Direct multi-step estimation and forecasting. Journal of Economic Surveys, 21, 746-785.

Cimadomo J. (2003). The effects of systematic monetary policy on sectors: A factor model analysis. ECARES - Universite Libre de Bruxelles, mimeo.

Clavel L. and Minodier C. (2009). A monthly indicator for the French business climate. Document de Travail No. G2009/02, Insee.

Clements M. and Galvao A.-B. (2008), Macroeconomic forecasting with mixed-frequency data : Forecasting output growth in the United States, Journal of Business and Economic Statistics, 26.

Connor G. and Korajczyk R.A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of Financial Economics, 15, 373-394.

Connor G. and Korajczyk R.A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. Journal of Financial Economics, 21, 255–289.

Connor G. and Korajczyk R.A. (1993). A test for the number of factors in an approximate factor model. Journal of Finance, 48, 1263–91.

Cornec M. and Deperraz, T. (2006). Un nouvel indicateur synthétique mensuel résumant le climat des affaires dans les services en France. Économie et Statistique, 395-396, 13-38.

Cornec M. (2006). Analyse factorielle dynamique multifréquence appliquée à la datation de la conjoncture française. Économie et Prévision, 2006/1, 172, 29-43.

Cristadoro R., Forni M., Reichlin L. and Veronese G. (2005). A Core Inflation Index for the Euro Area. Journal of Money, Credit and Banking, 37, 3, 539-560.

D'Agostino, A. and Giannone D. (2012). Comparing alternative predictors based on large-panel factor models. Oxford Bulletin of Economics and Statistics, 74, 2, 306-326.

Darné O. and Ferrara L. (2011). Identification of slowdowns and accelerations for the euro area economy. Oxford Bulletin of Economics and Statistics, 73, 3, 335-364.

Del Negro M. and Otrok C. (2007). 99 luftballons: Monetary policy and the house price boom across US states. Journal of Monetary Economics, 54, 1962-1985.

Del Negro M. and Otrok C. (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Technical report, Federal Reserve Bank of New York.

De Mol C., Giannone D. and Reichlin L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?, Journal of Econometrics, 146, 318-328.

Dempster A., Laird N. and Rubin D. (1977). Maximum likelihood estimation from incomplete data. Journal of the Royal Statistical Society, 14:1-38.

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

Ding A.A. and Hwang J.T.G. (1999). Prediction intervals, factor analysis models, and high-dimensional empirical linear prediction. Journal of the American Statistical Association, 94, 446–455.

Doz C. and Lenglart F. (1999). Analyse factorielle dynamique : test du nombre de facteurs, estimation et application à l'enquête de conjoncture dans l'industrie. Annales d'Économie et Statistiques, 54, 91 -127.

Doz C., Giannone D. and Reichlin L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics, 164, 188-205.

Doz C., Giannone D. and Reichlin L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics, 94, 1014-1024.

Eickmeier S. (2005). Common stationary and non-stationary factors in the euro area analyzed in a large-scale factor model. Discussion Paper No. 2/2005, Bundesbank.

Eickmeier, S. (2007). Business cycle transmission from the US to Germany - a structural factor approach. European Economic Review, 51, 521-551.

Eickmeier S. and Ziegler C. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. Journal of Forecasting, 27, 3, 237-265.

Eickmeier S., Lemke, W. and Marcellino, M. (2015). Classical time-varying FAVAR models - Estimation, forecasting and structural analysis. Journal of the Royal Statistical Society: Series A, 178, 493-533.

Eklund J. and Kapetanios G. (2008). A review of forecasting techniques for large data sets. Working Paper No 625, Department of Economics, Queen Mary, University of London.

Favero C., Marcellino M. and Neglia F. (2005). Principal components at work: The empirical analysis of monetary policy with large datasets. Journal of Applied Econometrics, 20, 603-620.

Ferrara L. and Marsilli C. (2013). Financial variables as leading indicators of GDP growth: Evidence from a MIDAS approach during the Great Recession. Applied Economics Letters, 20, 3, 233-237.

Forni M. and Lippi M. (1997). *Aggregation and the Microfoundations of Dynamic Macroeconomics*. Oxford University Press.

Forni M. and Lippi M. (2011). The general dynamic factor model: One-sided representation results. Journal of Econometrics, 163, 23-28.

Forni M. and Reichlin L. (1998). Let's get real: A factor-analytic approach to disaggregated business cycle dynamics. Review of Economic Studies, 65, 453–473.

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

Forni M., Hallin M., Lippi M. and Reichlin L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? Journal of Monetary Economics, 50, 1243-1255.

Forni M., Hallin M., Lippi M. and Reichlin L. (2004). The generalized factor model: Consistency and rates. Journal of Econometrics, 119, 231–255.

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

Forni M., Giannone D., Lippi M. and Reichlin L. (2009). Opening the black box: Structural factor models with large cross-sections. Econometric Theory, 25, 1319–1347.

Frale C., Marcellino M., Mazzi G.L. 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, 2, 439-470.

Geweke J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio Economic Models*, Aigner D.J. and Goldberger A.S. (eds). North Holland, Amsterdam.

Ghysels, E., Sinko and Valkanov (2007). MIDAS Regressions: Further results and new directions. Econometric Reviews, 26, 53-90.

Giannone D., Reichlin L. and Sala L. (2004). Monetary policy in real time. Dans Gertler M. and Rogoff K. (eds.) NBER Macroeconomics Annual, MIT Press.

Giannone D., Reichlin L. and Sala L. (2006). VARs factor models and the empirical validation of equilibrium of business cycle models. Journal of Econometrics, 132, 257-279.

Giannone D., Reichlin L. and Small D. (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 55, 665-676.

Gorman W. M. (1981). Some Engel curves. In Deaton A. (ed.) *Essays in the Theory and Measurement of Consumer Behavior in Honor of Sir Richard Stone*. Cambridge University Press, New York.

Hallin M. and Liska R. (2007). Determining the number of factors in the general dynamic factor model. Journal of the American Statistical Association, 102, 603-617.

Hallin M. and Liska R. (2011). Dynamic factors in the presence of blocks. Journal of Econometrics, 163, 29–41.

Hamilton J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57, 2, 357-384.

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

Harding M.C. (2009). Structural estimation of high-dimensional factor models. Mimeo, Department of Economics, Stanford University.

Helbling, T. and Bayoumi, T. (2003). Are they all in the same boat? The 2000-2001 growth slowdown and the G7-business cycle linkages. Working Paper, WP/03/46, IMF.

Jacobs J.P.A.M. and Otter P.W. (2008). Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach. Econometric Reviews, 27, 385–397.

Jones C.S. (2001). Extracting factors from heteroskedastic asset returns. Journal of Financial Economics, 62, 293–325.

Jungbacker B. and Koopman S.J. (2008). Likelihood-based analysis for dynamic factor models. Discussion Paper NO 2008-0007/4, Tinbergen Institute, University Amsterdam.

Jungbacker B., Koopman S.J. and van der Wel M. (2011). Maximum likelihood estimation for dynamic factor models with missing data. Journal of Economic Dynamics and Control, 35, 1358–1368.

Kapetanios G. (2004). A note on modelling core inflation for the UK using a new dynamic factor estimation method and a large disaggregated price index dataset. Economics Letters, 85, 63-69.

Kapetanios G. (2010). An alternative method for determining the number of factors in factor models with large datasets. Journal of Business and Economic Statistics, 28, 397-409.

Kapetanios G. and Marcellino M. (2004). A comparison of estimation methods for dynamic factor models of large dimension. Queen Mary Working Paper No. 489, London.

Kim C-J. and 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.

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

Kose M. A., Otrok C. and Whiteman C. (2003). International business cycles: World, region and country-specific factors. American Economic Review, 93, 4, 1216-1239.

Kose M. A., Otrok C. and Whiteman C. (2008). Understanding the evolution of world business cycles. Journal of International Economics, 75, 110-130.

Krolzig H.M. and Hendry D. (2001). Computer automation of general-to-specific model selection procedures, Journal of Economic Dynamics and Control, 25, 831-866.

Lawley D.N. and Maxwell A.E. (1971). *Factor Analysis in a Statistical Method* Butterworth, London.

Lewbel A. (1991). The rank of demand systems: Theory and nonparametric estimation. Econometrica, 59, 711-730.

Lopes H. F. and West, M. (2004). Bayesian model assessment in factor analysis. Statistica Sinica, 14, 41-67.

Mansour M.J. (2003). Do national business cycles have an international origin? Empirical Economics, 28, 223-247.

Marcellino M., Stock, J. and Watson M. (2005) Macroeconomic forecasting in the euro area: Country specific versus euro wide information. European Economic Review, 47, 1-18.

Marcellino M. and 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, 518-550.

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

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.

Matheson, T. (2014). New indicators for tracking growth in real time. Journal of Business Cycle Measurement and Analysis, 7, 51-71.

Motta G., Hafner C. and von Sachs R. (2011). Locally stationary factor models: Identification and nonparametric estimation. Econometric Theory, 27, 6, 1279-1319.

Mumtaz, H. and Surico, P. (2009). The transmission of international shocks: A factor-augmented VAR approach. Journal of Money, Credit and Banking, 41, 71-100.

Newbold and Harvey, A. (2002). Forecast combination and encompassing. In *A Companion to Economic Forecasting*, Chapter 12, Clements M. and Hendry D. (eds), Basil Blackwell, Oxford.

Nguiffo-Boyom M. (2006). Un indicateur de retournement conjoncturel pour la France: une application du modèle à facteur avec changements de régimes. Économie et Prévision, 2006/1, 172.

Onatski A. (2009). Testing hypotheses about the number of factors in large factor models. Econometrica, 77, 1447-1479.

Onatski A. (2010). Determining the number of factors from empirical distribution of eigenvalues. The Review of Economics and Statistics, 92, 4, 1004-1016.

Pẽna D. and Poncela, P. (2006a). Forecasting with non-stationary dynamic factor models. Journal of Econometrics, 119, 291-321.

Pẽna D. and Poncela, P. (2006b). Non-stationary dynamic factor analysis. Journal of Statistical Planning and Inference, 136, 1237-1257.

Ross S. (1976). The arbitrage theory of asset pricing. Journal of Economic Theory, 13, 341-360.

Ruenstler G., Barhoumi K., Cristadoro R., Reijer D., Jakaitiene A., Jelonek P., Rua A., Ruth K., Benk S., and Van Nieuwenhuyze C. (2009). Short-term forecasting of GDP using large monthly data sets: A pseudo real-time forecast evaluation exercise. Journal of Forecasting, 28, 7, 595-611.

Sargent T. and Sims C. (1977). Business cycle modelling without pretending to have too much a priori economic theory. In Sims C. (ed.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Mi nneapolis, Minneapolis.

Schumacher C. (2007). Forecasting German GDP using alternative factor models based on large datasets. Journal of Forecasting, 26, 271-302.

Schumacher C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. Economics Letters, 107, 2, 95-98.

Schumacher C. and 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, 368-398.

Shumway R.H. and Stoffer D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis, 3, 253–264.

Steiger J.H. (1979). Factor indeterminacy in the 1930s and the 1970s some interesting parallels. Psychometrika, 40, 157–167.

Stock J. and Watson M. (1989). New Indexes of Coincident and Leading Economic Indicators. In NBER Macroeconomics Annual, NBER.

Stock J. and Watson M. (1998). Diffusion indexes. Working Paper No 6702, NBER.

Stock J. and Watson M. (1999). Forecasting inflation. Journal of Monetary Economics, 44, 293–335.

Stock J. and Watson M. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics, 20, 147-162.

Stock J. and Watson M. (2005a). Implications of dynamic factor models for VAR analysis. Working Paper No 11467, NBER.

Stock J. and Watson M. (2005b). Understanding changes in international business cycle dynamics. Journal of the European Economic Association, 3, 5, 968-1006.

Stock J. and Watson M. (2006). Forecasting with many predictors. In Elliott G., Granger C.W.J. and Timmermann A. (eds.), *Handbook of Economic Forecasting*, Vol. 1, North-Holland, Amsterdam.

Timmerman, A. (2006), Forecast combinations. In Elliott G., Granger C.W.J. and Timmermann A. (eds.), *Handbook of Economic Forecasting*, Vol. 1, 135-196, North-Holland, Amsterdam.

Van Nieuwenhuyze C. (2006). A generalized dynamic model for the Belgian Economy - useful business cycle indicators and GDP growth forecasts. Working Paper No 80, National Bank of Belgium.

Watson M.W. and Engle R.F. (1983). Alternative algorithms for estimation of dynamic MIMIC, factor, and time varying coefficient regression models. Journal of Econometrics, 23, 385-400.

White H. (1982) Maximum likelihood estimation of misspecified models. Econometrica, 50, 1-25.

Yao Q. and Pan J. (2008). Modelling multiple time series via common factors. Biometrika, 95, 365-379.

11

Forecasting with Rich Data: Model Comparison and Evidence from European Countries

Authors

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

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Katerina Petrova
University of St. Andrews
Katerina.Petrova@st-andrews.ac.uk

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

11.1 Introduction

It has been a standard assumption in macroeconomic modeling that agents are processing all the available quantities of information when forming their expectations for the future. Also, policymakers traditionally have looked at a vast array of indicator series in the run-up to major policy decisions, or in the words of Lars Svensson (Svensson (2005)) about what central bankers do in practice: '(l)arge amounts of data about the state of the economy and the rest of the world ... are collected, processed, and analyzed before each major decision.' However, most traditional macroeconomic prediction approaches rarely consists of models that handle more than 10 variables, because it is either inefficient or downright impossible to incorporate a much larger number of variables in a single forecasting model and estimate it using standard econometric techniques. This failure of traditional macroeconomic forecasting methods prompted a new strand of research devoted to the theory and practice of alternative macroeconomic forecasting methods that utilize large data sets.

These alternative methods can be distinguished into two main categories. As, e.g., outlined in Hendry (1995), the methods of the first category involve inherently two steps: In the first step some form of variable selection is undertaken. The variables that are chosen are then used in a standard forecasting model. Recent developments in this line of research has focussed on automated model selection procedures in order to be better able to select the optimal predictors from large data sets; see Krolzig and Hendry (2001). An alternative group of forecasting methods consists of estimation strategies that allow estimation of a single equation model that utilizes *all* the information in a large data set and not just an 'optimal' subset of the available predictor series. This is a diverse group of forecasting methods ranging from factor-based methods to Bayesian regression and forecast combination. These two groups of methods inevitably overlap. However, we feel that the step of variable selection is, and involves methods that are, sufficiently distinct to merit separate mention and treatment. Instead, we focus in this paper on the latter group of data-rich forecasting methods.

Within the group of data-rich forecasting techniques, factor methods have gained a prominent place. These methods are related to the strict factor models used in finance, but, starting with Chamberlain and Rothschild (1983), they use weaker assumptions regarding the behavior of the idiosyncratic components, which allows the use of principal components in very large data sets to identify the common factors in such a data set. Stock and Watson (2002a) and Bai (2003) further formalized the underlying asymptotic theory. Stock and Watson (2002b) proved to be the starting point of a large empirical research output where, with mixed success, a limited number of principal components extracted from a large data set are used to forecast key macroeconomic variables.

However, the use of principal components does not always guarantee that the information extracted from a large number of predictors is useful for forecasting. Boivin and Ng (2006) make it clear that if the forecasting power comes from a certain factor, this factor can be dominated by other factors in a large data set, as the principal components solely provide the best fit for the large data set and not for the target variable. This could explain why in some empirical applications principal components (PC) factor models are dominated by Bayesian regression and forecast combinations. Under Bayesian regression one essentially estimates a multivariate regression consisting of *all* predictor variables, but with the regression coefficients shrunken to a value close to zero. Starting with Bates and Granger (1969), forecast combination involves the use of subsets of predictor variables in distinct forecasting models and the production of multiple forecasts for the target variable, which are then averaged to produce a final forecast.

The distinctive feature of these latter two approaches is that the information in a large data set is compressed such that this has explanatory power for the target variable. Note, however, that from an econometric perspective forecast combinations are *ad hoc* in nature, whereas it has been shown in De Mol et al. (2008) that Bayesian regression is theoretically related to PC-based factor models. Another technique that has potential for providing a solution to the issue of handling large dataset is partial least squares (PLS) regression. Although less widely known, PLS regression is an alternative data-rich approach that can be used for macroeconomic forecasting using very large data sets, irrespective of whether such a data set has a strong factor structure or not. PLS regression is implemented for large data sets through the construction of linear, or-

thogonal combinations of the predictor variables such that the linear combinations maximize the covariance between the target forecast variable and each of the common components constructed from the predictor variables. Although similar in spirit to PC regression, the explicit consideration of the target forecast variable addresses a major existing criticism towards PC regression as a forecasting technique.

For analysing the properties of the various data-rich methods, in particular PC, PLS and Bayesian regression, as well as a Monte Carlo exercise, we refer the reader to Groen and Kapetanios (2015). In this paper, we apply the methods and compare their forecasting performance using large datasets of various macroecomic, financial and survey data for several European counries.

The remainder of this paper has the following structure: Section 11.2 briefly discusses the different methods. Section 11.3 outlines the procedure for forecasting GDP growth and inflation in euro area countries. Section 11.4 concludes.

11.2 Methods for Data-Rich Macroeconomic Forecasting

A useful framework for studying existing methods is provided by the following general forecasting equation

$$y_t = \alpha' x_t + \epsilon_t; \quad t = 1, \ldots, T, \tag{11.1}$$

where y_t is the target of the forecasting exercise, $x_t = (x_{1t} \cdots x_{Nt})'$ is a vector of dimension $N \times 1$ and thus $\alpha = (\alpha_1 \cdots \alpha_N)'$ is also $N \times 1$. It is assumed that the number of indicator variables N is too large for α to be determined by standard methods such as ordinary least squares (OLS). The literature has proposed a number of ways how one can deal with this issue of large-dimensional data sets, of which we provide a selective review below. Before proceeding however, we need to stress that our assumed framework given by (11.1) is a significant deviation from, and, we argue, generalisation of, the existing literature which analyses the case where y_t does not depend on the large, observed dataset, x_t, but a small, unobserved set of variables, referred to as factors. We start our review with methods that have been shown to be applicable in this case. In each subsection we also present theoretical results and discussion on the properties of the methods under (11.1) while recaping the existing theory that relates to the old factor setup.

11.2.1 Factor Methods

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 forecasting equations which thereby becomes a standard forecasting equation as they only involve a few variables. The assumption is that the co-movements across the indicator variables can be captured by a $r \times 1$ vector of unobserved factors $F_t = (F_{1t} \cdots F_{rt})'$, i.e.

$$\tilde{x}_t = \Lambda' F_t + e_t \tag{11.2}$$

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 (11.2) 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 (11.1) $\alpha' = \tilde{\alpha}' \Lambda \tilde{x}_t$, 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. The main difference between different factor methods relate 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,b), 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 (PC) is simple. Estimates of Λ and the factors F_t are obtained by solving:

$$V(r) = \min_{\Lambda,F} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{x}_{it} - \lambda_i' F_t)^2, \tag{11.3}$$

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 (11.3) 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); Forni et al. (2004) among others) is designed to address this issue. 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)).

The setup whereby y_t depends on F_t, rather than directly on the observed x_t, has been dominant in the literature. However, it is reasonable to cast doubt on such a setup from a variety of directions. The first relates to the simplifying assumption implicit in (11.2) that for *any* potential target variable y_t one cannot find a better set of predictors than F_t. To see why such an assumption is not waranteed even within the confines of (11.2), it is instructive to consider allowing $y_t = x_{it}$ for some i. Denote by x_{-it} the subset of x_t that does not include x_{it}. Then, one can ask whether x_{-it} may contain information useful in forecasting x_{it}, that is not contained in F_t. To see that this is indeed the case we note that a standard assumption for factor models such as (11.2) allows for some correlation between elements of e_t. If that is the case, then there exist elements of x_{-it} whose idiosyncratic components are correlated with e_{it}. It then immediately follows that including these elements of x_{-it} as predictors in a predictive regression for x_{it} can result in a better specified predictive regression, with a smaller forecasting error, than the one obtained by simply using F_t as predictors. This casts doubt on the assumption that the generating model for y_t should be one that involves only F_t given by

$$y_t = \zeta' F_t + \epsilon_t, \tag{11.4}$$

with $\zeta = (\zeta_1 \cdots \zeta_r)'$ and r equals the number of unobserved factors that underlie the N predictor variables x_t.

To the best of our knowledge, the assumption in (11.4) has been made in all existing theoretical econometric work involving forecasting or modelling using large datasets, such as De Mol et al. (2008) or Kelly and Pruitt (2012). There is, at least, one further objection to this assumption. It is that, in fact, factors may not exist. While this is extreme and would imply that x_t are largely uncorrelated, which does not accord with empirical evidence, the degree to which factors exist has been discussed recently in the context of weak factor models. Perhaps, the most relevant work is that of Bailey et al. (2013) who analyse extensively such models and propose both a measure of the strength of factor models and an estimator for this measure. They also find that some datasets which have been used in the past to illustrate the use of strong factor models may in fact not follow such models. It is important to stress that this possibility has not been analysed extensively within the context of forecasting since the dominant assumption is a strong factor one, where all eigenvalues of $\Lambda'\Lambda$ in (11.2) are $O(N)$. It is then worth noting that the validity of PC as an estimator of the factors rests on relatively strong factor models. In particular, Kapetanios and Marcellino (2010) provide conditions for

consistency under weak factor model and show that PC is inconsistent if the factor model is too weak.

Having provided a justification for the use of (11.1), we note the properties of PC under (11.1). In particular, it is clear that the best forecast error that PC can achieve under (11.1) is $\alpha' e_t + \epsilon_t$ which clearly has larger variance than ϵ_t for all finite N. As $N \to \infty$, $\alpha' e_t + \epsilon_t$ has larger variance than ϵ_t if $\|\alpha\| = O(1)$.

11.2.2 Partial Least Squares Regression

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.[1] 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. The latter assumption has motivated the use of PC regression for macroeconomic forecasting, as the principal components under that assumption can under certain conditions identify the common factors in the data set (see, e.g., Bai (2003)). One of our contributions is, therefore, to develop the asymptotic properties of PLS regression under similar assumptions that $N,\ T \to \infty$ for data sets that have a common factor structure. In addition, we also investigate in that subsection the asymptotic properties of PLS regression when large data sets have a weak factor structure that possibly vanishes asymptotically.

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:
Algorithm 3.

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* $f_{j,t}$.
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 forecast by regressing y_t on $f_{j,t}$ $j = 1, ..., k$. Helland (1988, 1990) provide a general description of the partial least squares (PLS) regression problem. Helland (1988) shows that the estimates of the coefficients α in the regression of y_t on x_t, as in (11.1), obtained implicitly via PLS Algorithm 3 and a regression of y_t on $f_{j,t}$ $j = 1, ..., k$, are mathematically equivalent to

$$\hat{\alpha}_{PLS} = V_k (V_k' X' X V_k)^{-1} V_k' X' y \tag{11.5}$$

[1] Herman Wold and co-workers introduced PLS regression between 1975 and 1982, see, e.g., Wold (1982), whereas principal components regression as used by, e.g., Stock and Watson (2002a) has its roots in Pearson (1901). Since then it has received much attention in a variety of disciplines, especially in chemometrics, outside of economics.

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, (11.5) suggests that the PLS factors that result from Algorithm 3 span the Krylov subspace generated by $X'X$ and $X'y$, resulting in valid approximations of the covariance between y_t and x_t.

11.2.3 Bayesian (Shrinkage) Regression

Bayesian regression is a standard tool for providing inference for α in (11.1) and there exist 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 (11.1) given by

$$\hat{\alpha}_{BRR} = (X'X + vI)^{-1}X'y \tag{11.6}$$

where $X = (x_1, ..., x_T)'$, $y = (y_1, .., y_T)'$ and v is a shrinkage scalar parameter. The shrinkage estimator (11.6) 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 (11.6) 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 De Mol et al. (2008). De Mol et al. (2008) produce a consistency result forecasts produced using the above implementation of Bayesian regression.

11.2.4 A Comparison of the Methods

Regarding PLS regression, we note that Garthwaite (1994) provides a rationale to cast (*ad hoc*) forecast combinations in terms of the above described PLS framework. Essentially what Garthwaite (1994) shows is that a general PLS algorithm like Algorithm 3 can be expressed in terms of sequences of univariate regressions, where we regress u_t on $v_{i,t}$, $i = 1, ..., N$ and denote the OLS estimate of the coefficient of each regression by β_i. The j-th PLS factor then equals weighted average of $\beta_i v_{it}$: $f_{j,t} = \tilde{w}_j' v_t$ with $\tilde{w}_j = ((\beta_1 w_{1j}) \cdots (\beta_N w_{Nj}))'$ where $(w_{1j} \cdots w_{Nj})$ are given. Therefore, if one sets $(w_{1j} \cdots w_{Nj}) = (Var(v_{1t}) \cdots Var(v_{Nt}))$ Algorithm 3 follows, but if one assumes $(w_{1j} \cdots w_{Nj}) = (\frac{1}{N} \cdots \frac{1}{N})$ than in the *one-factor* case the Capistrán and Timmermann (2008) projection on equal-weighted mean (PEW) forecast combination approach follows. Thus, forecast combinations can be interpreted as restricted approximations to a one-factor PLS regression, with alternative specifications for $(w_{1j} \cdots w_{Nj})$ and often with *zero* intercept and slope coefficients in the final forecast regression.[2] By interpreting forecast combinations as a form of PLS regression, we obtain the underpinning for the relatively good performance of forecast combinations *vis-à-vis* PC regressions within different data environments.[3] Note, though, that PLS is much more general and it allows for several factors to be included in the forecast regression. In addition, De Mol et al. (2008) prove the existence of a form of asymptotic equivalence between PC regression and Bayesian regression when the underlying data comply with a dominant factor structure in (11.4). Thus, given such a dominant factor structure, Bayesian regression should be asymptotically equivalent to PLS regression and, under the one-factor assumption, forecast combinations.

[2] Granger and Ramanathan (1984) suggest regressing individual forecasts on the target variable with the resulting parameter estimates serving as combination weights. More generally, by specifying a loss function one can derive combination weights that are optimal under that specific loss function (Elliott and Timmermann (2004)). Timmermann (2006) provides a comprehensive survey of the forecast combination literature.

[3] For example, Faust and Wright (2009) show that forecast combination methods provide better out-of-sample performance than factor methods when applied to high-dimensional panels of U.S. macroeconomic data.

Therefore, the introduction of the PLS regression framework provide a means to asymptotically tie together different existing data-rich forecasting methods when a dominant factor structure is assumed to be present in the data. However, when there is a much weaker factor assumption, Groen and Kapetanios (2015) show that these asymptotic links no longer seem to hold, except for the link between PLS and forecast combinations, but in this case PLS regression nonetheless retains its desirable properties. However, it is reasonable to expect that in the case of weak factor models, Bayesian regression will share many of the attractive features of PLS and would dominate PC, under (11.1). As noted earlier, (11.1) has not been analysed for Bayesian regression, neither under strong nor under weak factor models, since the alternative model given by (11.4) has been analysed instead, even though it has a number of unattractive characteristics compared to (11.1), as discussed in Section 11.2.1.

11.3 Forecasting GDP growth and Inflation in Euro Countries

11.3.1 Data

In order to better understand the strengths and weaknesses of the approaches, we design a pseudo out-of-sample forecasting exercise. We apply the models outlined in the previous section to Eurozone aggregate data and to several European countries; namely, UK, Germany, France and Italy. In particular, we estimate the three models for each country, using a medium and a small data set. The variables which we are interested in forecasting are quarterly GDP growth and anual CPI inflation. For the medium sized models, we have on average around 12 variables per country; those include economic activity variables, such as unemployment rate, industrial production; leading indicators; financial variables, such as stock market indices; interest rates and spreads; price indices and also survey data, such as sentiment indicators. The large models contain between 57 (for France) and 128 (for the UK) variables. In addition to the national data, the large datasets also include some international series such as commodity and oil prices as well as some US series. Detailed desciption of the dataset including sample sizes and forecast origins for each country can be found in the Appendix.

Due to the heterogeneity in number of observations, evaluation periods and the nature of the data available for each country, the results in terms of absolute forecast errors, for instance, are not be directly comparable across countries. For example, the Eurozone aggregate data has a relatively short span; with an out-of-sample period 2003Q1-2013Q3; hence, forecasts errors will be inflated with the financial crisis so in absolute sense these are much larger than for instance the forecast errors for the UK, which has an out-of-sample evaluation period 1993Q1-2013Q3.

11.3.2 Models

We generate forecasts 1 to 8 quarters ahead. Our forecasting scheme is direct with an expanding in-sample. In particular, for the factor models we have

$$y_{t+h} = \alpha + \beta_1 y_t + \beta_2 y_{t-1} + ... + \beta_p y_{t-p+1} + \gamma_1 f_{t,1} + \gamma_2 f_{t,2} + ... + \gamma_r f_{t,r} + \varepsilon_{t+h}$$

where α is a constant, p is the optimal lag for y_t, obtained by Bayesian Information Criterion, f_t are the factors, obtained either by Principle Component or through the Partial Least Square Procedure, and r is the optimal number of factors, obtained using a cross-validation procedure, in order to minimise the forecast errors in an in-sample evaluation period.

The forecast at horizon h is given by

$$\hat{y}_{T+h} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} + ... + \hat{\beta}_p y_{T-p+1} + \hat{\gamma}_1 f_{T,1} + \hat{\gamma}_2 f_{T,2} + ... + \hat{\gamma}_r f_{T,r}$$

and to obtain density forecasts, we employ wild bootstrap simulating from the residuals $\hat{\varepsilon}_{t+h}$.

Similarly, the Bayesian shrinkage model is given by:

$$y_{t+h} = \alpha + \beta_1 y_t + \beta_2 y_{t-1} + ... + \beta_p y_{t-p+1} + \zeta_1 x_{t,1} + \zeta_2 x_{t,2} + ... + \zeta_n x_{t,n} + \varepsilon_{t+h}$$

where x_t are the variables from our dataset directly. The ζ's are estimated using the Bayesian shrinkage procedure, outlined in the previous section; with a shrinkage parameter, chosen optimally as in the factor models, minimising the forecast errors in an in-sample evaluation period. The forecast h steps ahead is

$$\hat{y}_{T+h} = \hat{\alpha} + \hat{\beta}_1 y_T + \hat{\beta}_2 y_{T-1} + ... + \hat{\beta}_p y_{T-p+1} + \hat{\zeta}_1 x_{T,1} + \hat{\zeta}_2 x_{T,2} + ... + \hat{\zeta}_n x_{T,n}.$$

In addition, we estimate a AR (p) model, as our univariate benchmark:

$$y_{t+h} = \alpha + \beta_1 y_t + \beta_2 y_{t-1} + ... + \beta_p y_{t-p+1}$$

11.3.3 Results

The RMSFEs and log predictive scores are reported in the tables in the Appendix and are relative to the benchmark AR (p) model. The numbers reported for the RMSFEs are in ratios, so that numbers smaller than one for the RMSFEs. For the log scores, the values in the tables are differences from the benchmark so positive values indicate better performance over the benchmark.

For the Eurozone, both Principle Component (PC) and Partial Least Squares (PLS) perform similarly for GDP growth and deliver gains over the univariate model at one step ahead. The Bayesian shrinkage model, on the other hand, seems to perform well at longer horizons. The forecasts for inflation suggest a different ranking, with PLS performing worse than the PC at short and medium horizons and Bayesian shrinkage again delivering gains at medium horizons. There are mixed results for the density forecasts for GDP, but both factor models deliver small improvements in terms of log predictive score for GDP at one step ahead and for inflation at medium to long horizons.

Adding more data leads to better forecast performance for GDP for PC at one step ahead while the gains for inflation remain even at longer horizons. This is expected, as the large datasets contain a variety of financial and survey series which have been documented to have a good predictive power especially for the short run. PLS's performance, on the other hand deteriorates.

Furthermore, it is clear that all models deliver forecast uncertainty in terms of density forecast which is not wide enough, implied by the U-shaped Probability Integral Transformations (PITs). When adding more data, those PITs become even less uniform and this is expected, as additional information reduces uncertainty. Moreover, the poor density performance might be a consequence of choosing optimally the number of factors and shrinkage parameter in order to optimise point and not density forecasts.

For France, the Bayesian shrinkage model delivers the best GDP forecast and performance improves further when more data is added. For example, at one step ahead, it delivers 12% more accurate forecast than the univarite model. For inflation, there are insignificant gains over the univariate model using medium datasets, and increasing the size of the data, does not improve performance.

When forecasting Italy's GDP, we find again that the Bayesian Shrinkage model performs best with considerable and statistically significant improvements over the benchmark. For inflation, PLS seems to do better at short horizons while Bayesian shrinkage delivers improvements at long horizons. Adding more data impoves all models' performance for GDP at one step ahead and Bayesian shrinkage remains the best model for GDP and inflation at long horizons. Density performance improves for the two factor models when larger datasets are used. An interesting point emerges from the PITs for GDP, where both the medium and large models at one step ahead deliver asymmetric PITs, implying that all models deliver upward biased forecasts; this bias disappears for the large Bayesian shrinkage model when more data is added.

For Germany, among the medium models, Bayesian shrinkage is the best model for GDP with 11% statistically significant gains over the AR(p) at one step ahead, while there are virtually no gains over the AR (p) model for both Factor models. Surprisingly, adding more data deteriorates the performance of all models. The picture emerging from the inflation forecast is clear: we find evidence that inflation is hard to forecast with either approach even with rich macro data and the univariate model is difficult to outperform.

Finally, for the UK, the GDP forecasts with medium models do not improve over the benchmark. However, adding more data delivers marginal imrpovements for the Bayesian shrinkage for both point and density forecasts. The results for inflation also point to Bayesian shrinkage as the best performing model for both point and density forecasts especially at longer horizons.

11.4 Conclusion

To summarise, we compare Principle Component, Partial Least Squares and Bayesian Shrinkage models in an out-of-sample forecasting exercise on a set of European countries. All three methods are designed to deal with a large set of predictors and provide different solutions to the curse of dimensionality. The results we presented suggest several conclusions. First, all methods deliver gains for most countries at shorter horizons when forecasting GDP growth, implying that financial and macro data have a predictive power over economic activity. The Bayesian shrinkage model performs better than the two factor models. On the other hand, inflation is harder to predict and a univariate model which only features past dynamics of inflation itself is difficult to outperform. Second, adding more data improves further the forecasts for GDP and inflation for some countries while it deteriorates performance of others. Third, the results on density forecasts are mixed; one message that emerges from the probability integral transformations is that uncertainty about the forecasts is too small. This could be an outcome of the bootstrap procedure that we use and a potential solution could be to implement Fixed Regressor Bootstrap, which in addition features parameter uncertainty.

Annex

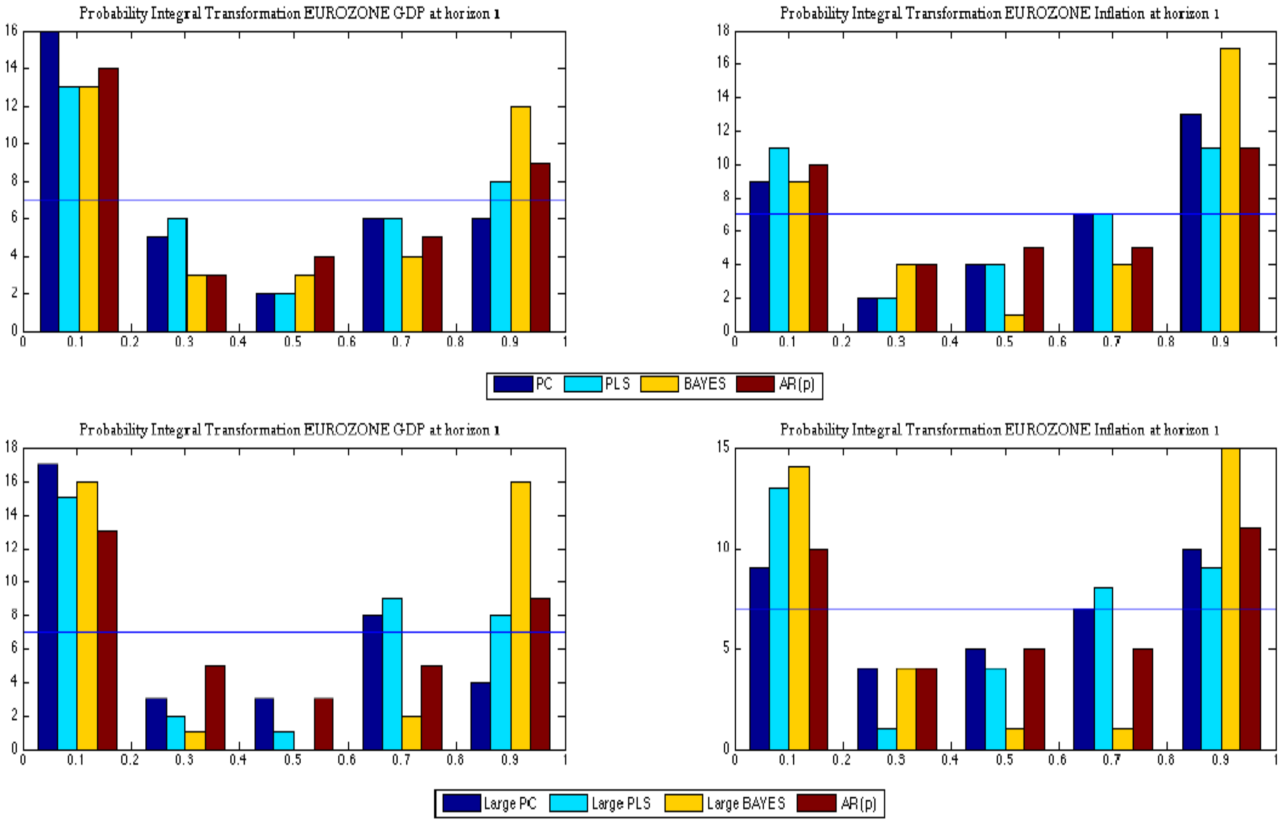
11.A Model Comparison Results

11.A.1 Eurozone

MODEL COMPARISON EUROZONE						
MEDIUM SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	0.86	0.9	1.08	0.98	1	1.3
2	1.1	1.17	0.89	0.93*	1.06	0.92
4	1.13	1.11	0.93	0.92	1.01	0.86
8	1.08	1.01	1.14	1.04	1.01	1.01
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	0.07	0.07	0.05	-0.05	-0.27	-0.61
2	-0.3	-0.38	-0.14	0.12	0.14	-0.04
4	-0.35	-0.27	-0.09	0.7	0.54	-0.03
8	-0.25	-0.02	-0.19	0.23	0.51	0.2
LARGE SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	0.78*	1.01	1.18	0.94	1.09	0.99
2	1.1	1	0.84	0.83*	1.1	0.76
4	1.01	1.03	0.73	0.88*	1.08	0.83
8	1.14	1.13	0.91	1.03	1	0.99
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.01	-0.16	-1.94	-0.01	-0.13	-0.43
2	-0.3	-0.18	-0.85	0.17	0.19	-0.32
4	-0.06	-0.03	0.18	0.22	-0.36	-0.03
8	-0.21	-0.04	-0.45	-0.08	0.53	-0.67

Note: RMSFEs entries are expressed as a ratio over the AR (p) benchmark, smaller than one numbers imply better performance of the particular model; Log Scores entries are expressed as a difference from the AR (p) benchmark, positive numbers imply better performance of the particular model. *, ** and *** imply significance at 90, 95 and 99 confidence level respectively of the Diebold Mariano test of equal performance.

Figure 11.A.1: Euro area

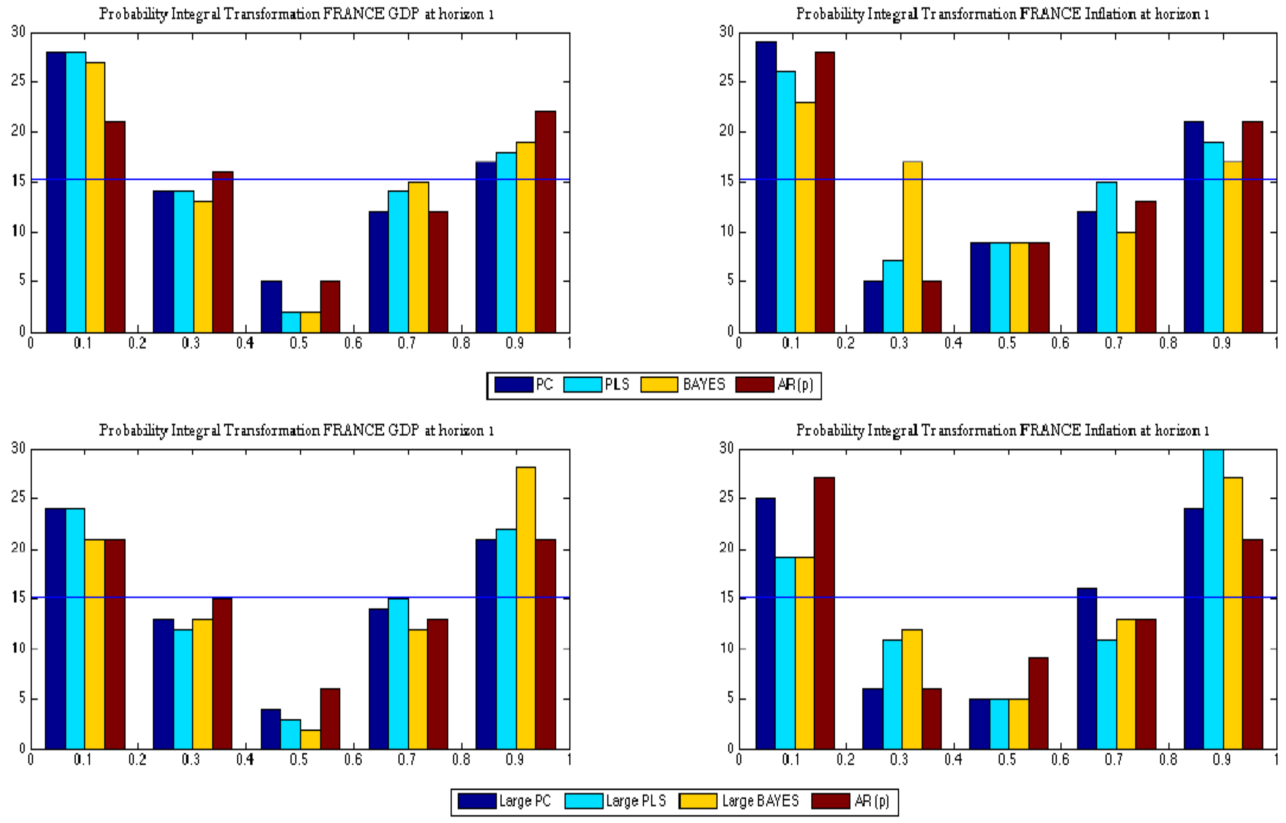


11.A.2 France

MODEL COMPARISON FRANCE						
MEDIUM SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	0.96*	1.01	0.90**	0.97	1	1.13
2	1.05	1.06	1.04	0.95	0.94	0.95
4	1.03	1.05	1.04	1	1.01	0.99
8	1	0.99	0.95*	1.06	1.11	1.09
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.01	-0.03	0.06	0.01	0.01	0.04
2	-0.09	-0.1	-0.14	0.02	0.01	-0.01
4	-0.03	-0.09	-0.15	-0.04	-0.16	0.01
8	-0.03	-0.02	0.02	-0.12	-0.24	-0.09
LARGE SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	1.02	0.99	0.88**	1.03	1.06	1.22
2	1.05	1.08	0.95	1.06	1.06	0.94
4	1	1.06	0.93*	1.1	1.07	0.92
8	1	0.99	0.96	1.13	1.08	1.02
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.02	-0.07	-0.03	-0.06	-0.14	-0.25
2	-0.11	-0.16	-0.11	-0.11	-0.06	-0.08
4	-0.08	-0.25	0.04	-0.3	-0.17	0.05
8	-0.01	-0.13	-0.06	-0.45	-0.23	-0.08

Note: RMSFEs entries are expressed as a ratio over the AR (p) benchmark, smaller than one numbers imply better performance of the particular model; Log Scores entries are expressed as a difference from the AR (p) benchmark, positive numbers imply better performance of the particular model. *, ** and *** imply significance at 90, 95 and 99 confidence level respectively of the Diebold Mariano test of equal performance.

Figure 11.A.2: France

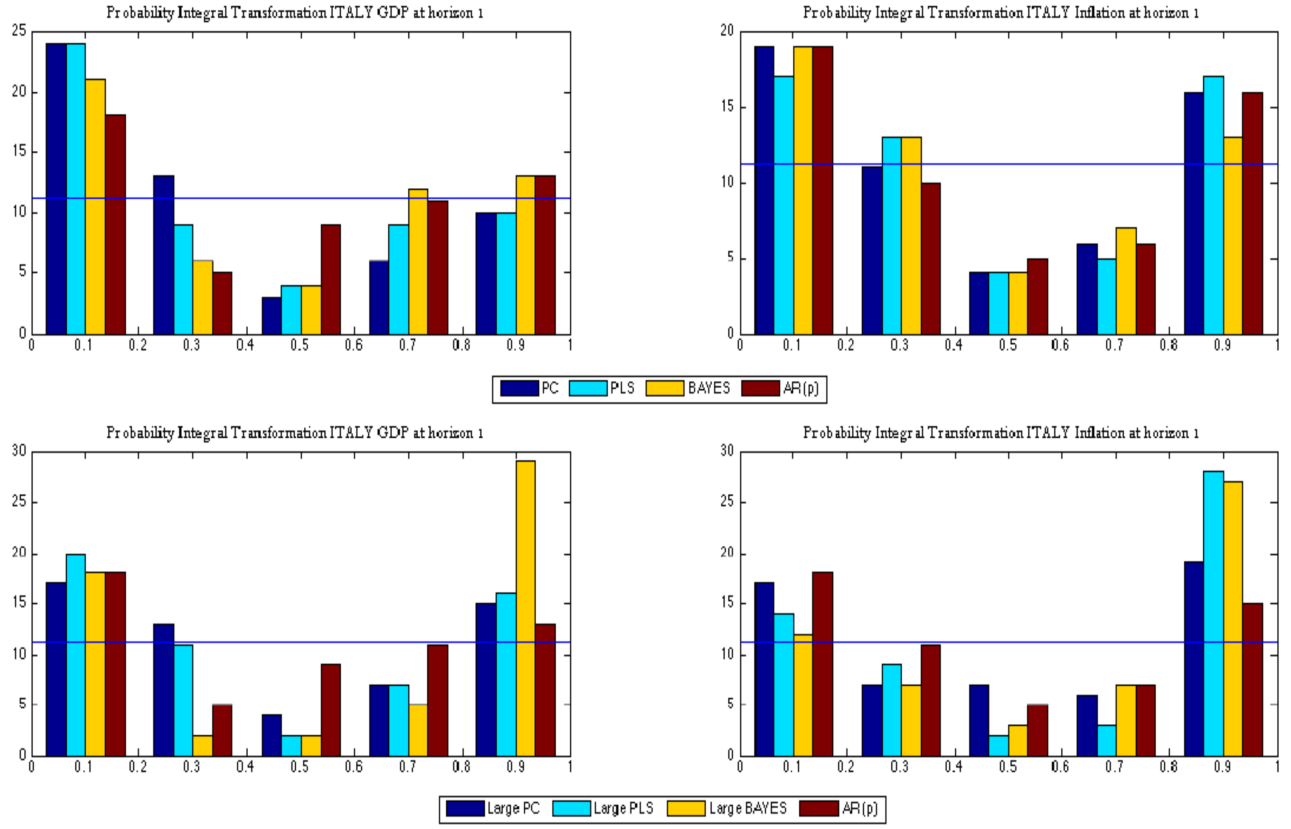


11.A.3 Italy

MODEL COMPARISON ITALY						
MEDIUM SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	1	0.96	0.88**	1.01	0.95*	1.29
2	1.03	1.05	0.94*	1.02	0.96	1.1
4	1.01	1.01	0.96	1.01	0.96	0.91*
8	0.99	0.97	0.96	1.18	1.11	0.94
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.36	-0.16	0.02	0	0.05	-0.33
2	-0.07	-0.34	-0.08	0.01	0.04	-0.18
4	0.01	-0.03	-0.03	0.05	0.05	0.17
8	-0.09	-0.04	0.06	-0.22	-0.18	0.06
LARGE SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	0.82	0.96	0.81	0.98	0.97	1.45
2	1.01	1.05	0.94	0.92	1.02	0.93
4	0.94	1	0.95	1.1	1.17	0.82*
8	0.98	1.01	0.94	1.02	1.07	0.89
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	0.13	-0.13	-0.65	-0.02	-0.11	-0.77
2	0.33	0.41	-0.01	-0.02	-0.29	-0.16
4	0	0.08	-0.26	-0.29	-0.35	0.08
8	-0.1	-0.08	-0.16	-0.09	-0.04	-0.1

Note: RMSFEs entries are expressed as a ratio over the AR (p) benchmark, smaller than one numbers imply better performance of the particular model; Log Scores entries are expressed as a difference from the AR (p) benchmark, positive numbers imply better performance of the particular model. *, ** and *** imply significance at 90, 95 and 99 confidence level respectively of the Diebold Mariano test of equal performance.

Figure 11.A.3: Italy

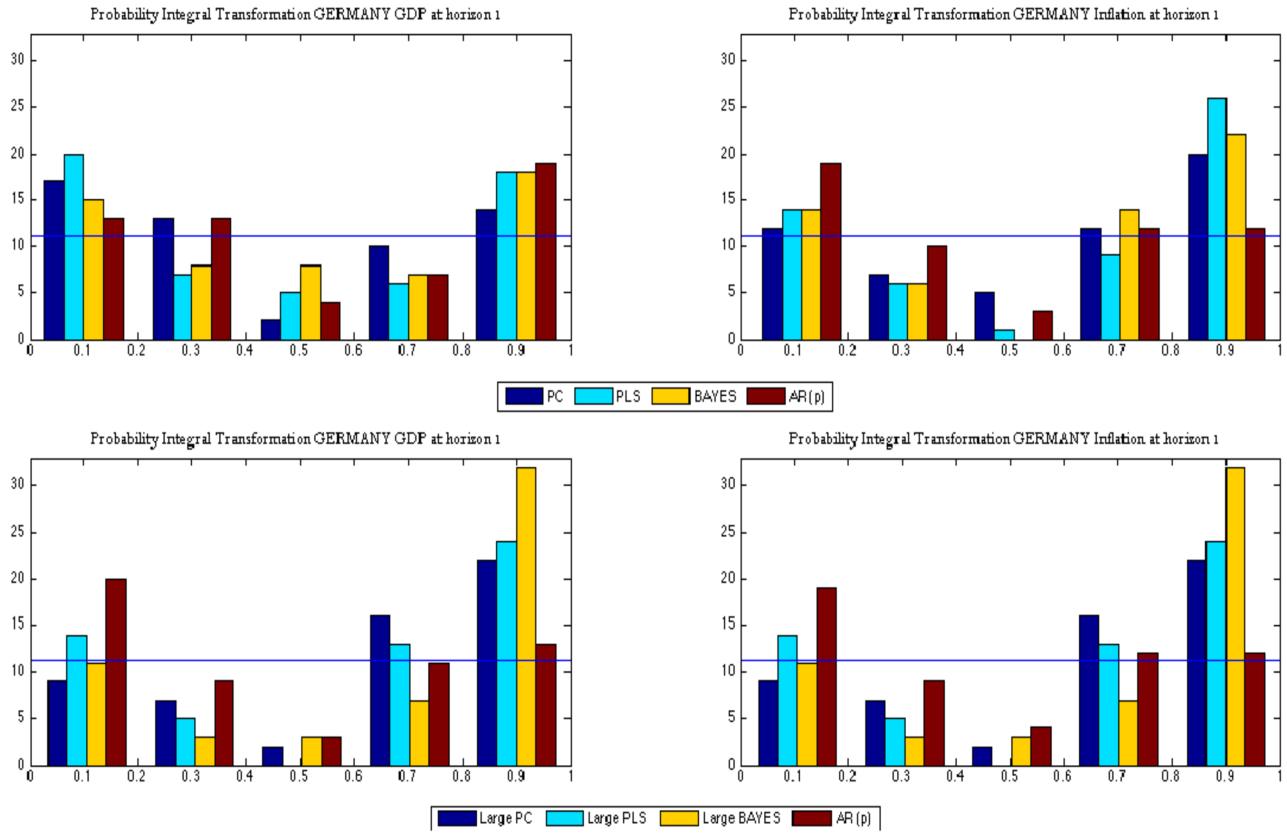


11.A.4 Germany

MODEL COMPARISON GERMANY						
MEDIUM SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	0.99	1.01	0.89**	1.15	1.18	1.13
2	1.01	1.07	0.97	1.22	1.27	1.06
4	1.03	1.03	0.99	1.25	1.45	0.9
8	1.03	1.06	1.04	1.22	1.41	1.08
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.06	-0.13	-0.05	-0.18	-0.37	-0.17
2	0.05	-0.07	0.08	-0.34	-0.55	-0.38
4	-0.12	-0.09	-0.01	-0.35	-0.73	0.02
8	-0.19	-0.25	-0.17	-0.31	-0.48	-0.11
LARGE SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	1.08	1.23	1.36	1.09	1.23	1.36
2	1.21	1.47	1.12	1.21	1.47	1.12
4	1.37	1.75	0.93	1.37	1.75	0.93
8	1.61	1.92	1.13	1.61	1.92	1.13
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.05	-0.37	-0.66	-0.08	-0.35	-0.71
2	-0.39	-0.76	-0.87	-0.42	-0.72	-0.9
4	-0.88	-0.98	-0.29	-0.85	-1.38	-0.32
8	-1.14	-1.39	-0.3	-1.16	-1.39	-0.34

Note: RMSFEs entries are expressed as a ratio over the AR (p) benchmark, smaller than one numbers imply better performance of the particular model; Log Scores entries are expressed as a difference from the AR (p) benchmark, positive numbers imply better performance of the particular model. *, ** and *** imply significance at 90, 95 and 99 confidence level respectively of the Diebold Mariano test of equal performance.

Figure 11.A.4: Germany

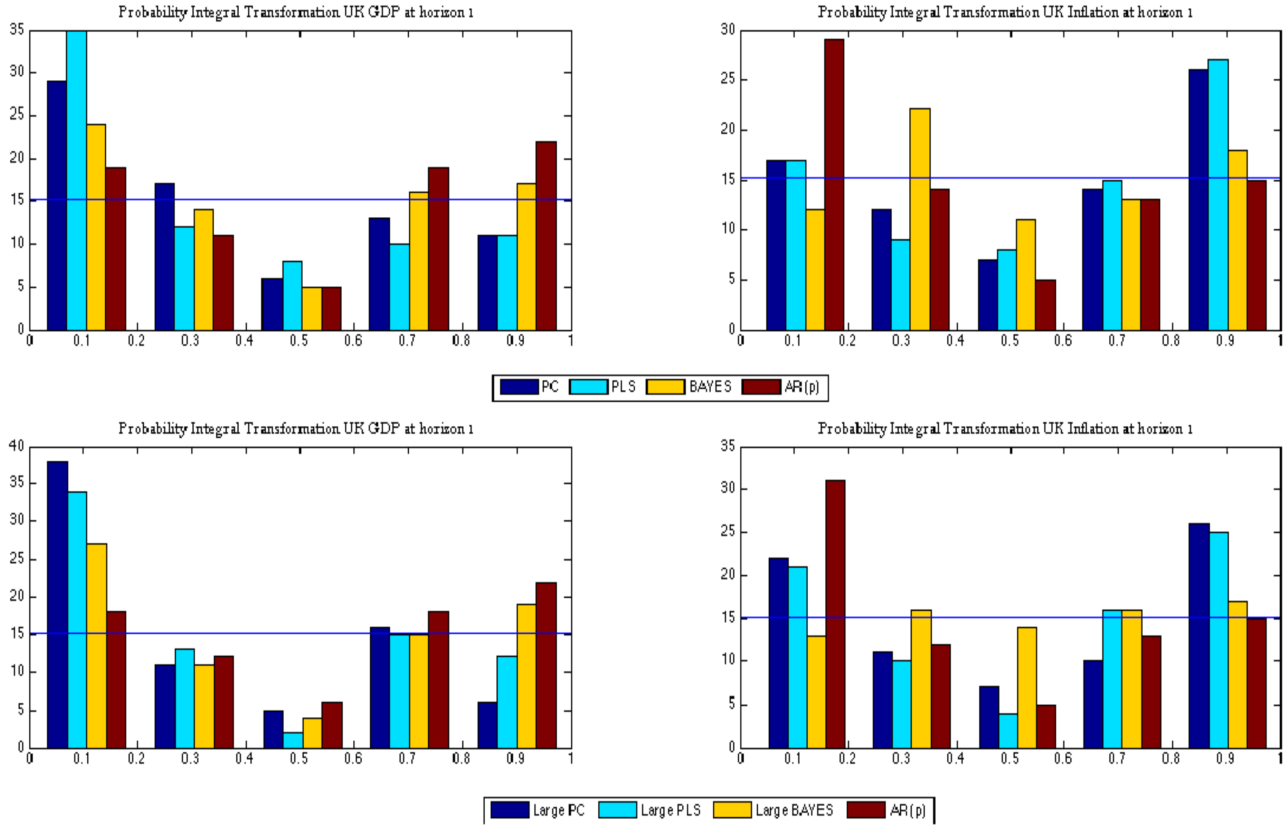


11.A.5 UK

MODEL COMPARISON UK						
MEDIUM SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	1.08	1.11	1.04	1.06	1.02	1.07
2	1.02	1.04	0.98	1.11	1.06	0.9
4	1.03	1.02	1.01	1.21	1.07	0.71**
8	1.07	1.08	1.01	1.06	1.12	0.78
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.07	-0.08	-0.01	0.03	0.03	0.09
2	0.05	0.03	0.12*	-0.04	-0.02	0.13
4	-0.02	-0.01	0	-0.09	0.02	0.46**
8	-0.04	-0.13	0	-0.14	-0.2	0.19
LARGE SIZE MODELS						
	Principle Component	Partial Least Squares	Bayesian Shrinkage	Principle Component	Partial Least Squares	Bayesian Shrinkage
Horizon	GDP RMSFEs relative to AR (p) Direct			Inflation RMSFEs relative to AR (p) Direct		
1	1.11	1.05	0.99	0.99	1.06	1.11
2	1.02	1.02	0.96	1	1.07	0.93
4	1.04	1.01	0.97	1.03	1.11	0.80*
8	1.1	1.1	0.97	1.09	0.87*	0.85
Horizon	GDP LPS relative to AR (p) Direct			Inflation LPS relative to AR (p) Direct		
1	-0.07	-0.11	-0.04	0.01	-0.08	0.05
2	0.01	0.01	0.06	0.03	-0.08	0.15*
4	-0.17	-0.03	0.01	-0.02	-0.1	0.30*
8	-0.09	-0.21	0.02	-0.33	0.11	-0.23

Note: RMSFEs entries are expressed as a ratio over the AR (p) benchmark, smaller than one numbers imply better performance of the particular model; Log Scores entries are expressed as a difference from the AR (p) benchmark, positive numbers imply better performance of the particular model. *, ** and *** imply significance at 90, 95 and 99 confidence level respectively of the Diebold Mariano test of equal performance.

Figure 11.A.5: United Kingdom



11.B Data Description

Country / Region	Sample Start	Sample End	Forecast Origin	Number of Forecasts
Eurozone	1998Q2	2013Q3	2003Q1	35
France	1983Q1	2013Q3	1993Q1	75
Italy	1991Q1	2013Q3	1998Q1	55
Germany	1991Q1	2013Q3	1998Q1	55
UK	1975Q1 / 1980Q1	2013Q3	1993Q1	75

11.B.1 Medium-Sized Models

EUROZONE	
EK INDUSTRIAL PRODUCTION: MANUFACTURING (EA18) VOLA	EKIPMAN.G
EK UNEMPLOYMENT (EA18) VOLA	EKESTUNPO
EK INDUSTRIAL PRODN. - CONSUMER NON DURABLES (%MOM) (EA18) SADJ	EKESICN%Q
EK NEW RESIDENTIAL BUILDINGS - COST INDEX (EA18) NADJ	EKECEIBCF
EURO STOXX - PRICE INDEX	DJEURST
BD DISCOUNT RATE / SHORT TERM EURO REPO RATE	BDPRATE.
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
EK ECONOMIC SENTIMENT INDICATOR (EA18) VOLA	EKEUSESIG
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
EK CONSTRUCTION SURVEY: EMPLOYMENT EXPECTATIONS (EA) SADJ	EK45.4BSQ
SPREAD GERMANY: BD LONG TERM GOVERNMENT BOND YIELD 9-10 YRS MINUS BD 3-M FIBOR	BDGBOND. - BDOIR076R
GERMANY	
BD INDUSTRIAL PRODUCTION INCLUDING CONSTRUCTION (CAL ADJ) VOLA	BDIPTOT.G
BD CNSTR.IND.: CAPACITY UTILIZATION SADJ	BDIFDCTNQ
BD UNEMPLOYMENT RATE, TOTAL SADJ	BDESUNEMO
BD EMPLOYMENT DURATION - SHORT-TERM WORKERS VOLN	BDEMPSTWP
BD NEW ORDERS TO MANUFACTURING - DOMESTIC: CONSUMER GOODS VOLA	BDDCNORDG
BD DAX SHARE PRICE INDEX, EP NADJ	BDSHRPRCF

BD REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	BDOCC011
BD CONSTRUCTION ORDERS RECEIVED - RESIDENTIAL CONSTRUCTION VOLA	BDHOUSE.G
BD DISCOUNT RATE / SHORT TERM EURO REPO RATE	BDPRATE.
BD PPI: INDL. PRODUCTS, TOTAL, SOLD ON THE DOMESTIC MARKET NADJ	BDPROPRCF
BD CONSUMER CONFIDENCE INDICATOR - GERMANY SADJ	BDCNFCONQ
BD HWWI IDX OF WORLD MKT.PRC.OF RAW MATS,EURO AREA: EXCL.ENERGY	BDIUW501F
SPREAD GERMANY: BD LONG TERM GOVERNMENT BOND YIELD 9-10 YRS MINUS BD 3-M FIBOR	BDGBOND. - BDOIR076R
FRANCE	
FR INDUSTRIAL PRODUCTION VOLA	FRIPTOT.G
FR UNEMPLOYMENT RATE, TOTAL SADJ	FRESUNEMO
FR UNEMPLOYMENT (HARMONIZED): TOTAL TRND	FRESQT8JT
FR BANQUE DE FRANCE SVY.: BUSINESS SENTIMENT INDICATOR(CAL ADJ)	FRSURCBSQ
FR SURVEY: MFG. OUTPUT - ORDER BOOK & FOREIGN DEMAND SADJ	FRSURFMPQ
FR SURVEY: MFG. OUTPUT - FINISHED GOODS INVENTORIES SADJ	FRSURSMPQ
FR SHARE PRICE INDEX - SBF 250 NADJ	FRSHRPRCF
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
FR SHARE PRICE INDEX - SBF 250 NADJ	FRSHRPRCF
FR REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	FROCC011
FR AVERAGE COST OF FUNDS FOR BANKS / EURO REPO RATE	FRPRATE.
SPREAD: FR GOVERNMENT BOND YIELD MINUS FR CAPITAL MRKT YIELDS-13-WEEK T BILLS	FRGBOND.- FRGBILL3
ITALY	
IT INDUSTRIAL PRODUCTION VOLA	ITIPTOT.G
IT UNEMPLOYMENT RATE, TOTAL SADJ	ITESUNEMO
IT UNEMPLOYMENT VOLA	ITESTUNPO
IT IND.: OVERALL - EMPL EXPECT SADJ	ITTTA7BSQ
IT NEW RESIDENTIAL BUILDINGS - COST INDEX NADJ	ITECEIBCF
IT DISCOUNT RATE / SHORT TERM EURO REPO RATE	ITPRATE.
IT MILAN COMIT GENERAL SHARE PRICE INDEX (EP) NADJ	ITSHRPRCF
IT REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	ITOCC011
IT PPI - LINKED & REBASED NADJ	ITPROPRAF
IT ECONOMIC SENTIMENT INDICATOR VOLA	ITEUSESIG
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
SPREAD: IT GOVERNMENT BOND GROSS YIELD MINUS ITALY T-BILL AUCT. GROSS 3 M RATE	ITGBOND.-ITBT03G
UK	
UK INDEX OF PRODUCTION - ALL PRODUCTION INDUSTRIES VOLA	UKIPTOT.G
UK LFS: IN EMP.: AGED 16+: ANNUAL = SPRING QUARTER(MAR-MAY) VOLA	UKMGRZ..O
UK UNEMPLOYMENT RATE SADJ	UKUN%TOTQ
UK GFK CONSUMER CONFIDENCE INDEX NADJ	UKGFKCCNR
UK LSL/ ACAD AVERAGE HOUSE PRICE CURA	UKFTHPI.B
UK FT ALL SHARE INDEX (EP) NADJ	UKSHRPRCF
UK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	UKOCC011
UK INTERBANK RATE - 3 MONTH (MONTH AVG)	UKINTER3
UK PPI - OUTPUT OF MANUFACTURED PRODUCTS (HOME SALES) NADJ	UKPROPRCF
UK BANK OF ENGLAND BASE RATE (EP)	UKPRATE.
UK YIELD 10-YEAR MINUS UK DISCOUNT RATE 3-MONTH TREASURY BILLS	UKOIR080R - UKOIR077R
UK RPI - ALL ITEMS EXCLUDING MORTGAGE INTEREST NADJ	UKRPAXMIF
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF

11.B.2 Large-Sized Models

EUROZONE	
BD LONG TERM GOVERNMENT BOND YIELD - 9-10 YEARS	BDGBOND.
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
BD DISCOUNT RATE / SHORT TERM EURO REPO RATE	BDPRATE.
CH EXPORTS CURN	CHEXPGDSA
CH INDUSTRIAL PRODUCTION INDEX VOLN	CHIPTOT.H
CH GOLD AND FOREIGN RESERVES - FOREIGN RESERVE CURN	CHRESERVA
EURO STOXX - PRICE INDEX	DJEURST
EK IND.: CMPTR, ELECTRONIC & OPTICAL PRDS - EXP ORD BOOKS SADJ	EK26.3.BQ
EK IND.: ELECTRICAL EQP - EXP ORD BOOKS SADJ	EK27.3.BQ
EK CNSTR.: TOT BLDG - EMPL EXPECT. SADJ	EK41.4.BQ
EK CNSTR.: TOT BLDG - PRICE EXPECT. SADJ	EK41.5.BQ
EK CONSTRUCTION SURVEY: ORDER BOOK POSITION (EA) SADJ	EK45.3BSQ
EK CONSTRUCTION SURVEY: EMPLOYMENT EXPECTATIONS (EA) SADJ	EK45.4BSQ
EK CONSUMER CONFIDENCE INDICATOR (EA) SADJ	EKCNFCONQ
EK IND.: CONSUMER GDS - PROD EXPECT SADJ	EKCNS5.BQ
EK IND.: CONSUMER GDS - SELL PRICE EXPECT SADJ	EKCNS6.BQ
EK IND.: CONSUMER GDS - EMPL EXPECT SADJ	EKCNS7.BQ
EK IND.: DURABLE CONSUMER GDS - EXP ORD BOOKS SADJ	EKCUR3.BQ
EK ECB HARMONISED COMPETITIVENESS INDICATOR - EURO AREA NADJ	EKECBHCIF
EK NEW RESIDENTIAL BUILDINGS - COST INDEX (EA18) NADJ	EKECEIBCF
EK INDUSTRIAL PRODUCTION - CONSTRUCTION (EA18) VOLA	EKESCONMG
EK CPI - ENERGY (%YOY) (EA18) NADJ	EKESCPEN%
EK CPI - FOOD INCLUDING ALCOHOL & TOBACCO (%YOY) (EA18) NADJ	EKESCPFT%
EK CPI - SERVICES EXCLUDING GOODS (EA18) NADJ	EKESCPSVF
EK CPI - TRANSPORT (%YOY) (EA18) NADJ	EKESCPTP%
EK CPI-HOUSING,WATER,ELECTRICITY,GAS & OTH.FUELS(%YOY)(EA18)	EKESCPUT%
EK INDUSTRIAL PRODN. - CONSUMER NON DURABLES (%MOM) (EA18) SADJ	EKESICN%Q
EK INDUSTRIAL PRODUCTION - CONSUMER DURABLES (%MOM) (EA18) SADJ	EKESICO%Q
EK INDUSTRIAL PRODUCTION - CAPITAL GOODS (%MOM) (EA18) SADJ	EKESICT%Q
EK INDUSTRIAL PRODUCTION - ENERGY (EA18) VOLA	EKESIENGG
EK INDUSTRIAL PRODUCTION - INTERMEDIATE GOODS (%MOM) (EA18) SADJ	EKESIIT%Q
EK UNEMPLOYMENT (EA18) VOLA	EKESTUNPO
EK INDUSTRY SURVEY: EXPORT ORDER-BOOK LEVELS (EA) SADJ	EKEUIMF3Q
EK INDUSTRY SURVEY: PRODUCTION EXPECTATIONS (EA) SADJ	EKEUIMF5Q
EK INDUSTRY SURVEY: EMPLOYMENT EXPECTATIONS (EA) SADJ	EKEUIMF7Q
EK CONSTRUCTION SURVEY: EMPLOYMENT EXPECTATIONS (EA) NADJ	EKEUSBEMR
EK CONSTRUCTION SURVEY: PRICE EXPECTATIONS (EA) NADJ	EKEUSBPRR
EK ECONOMIC SENTIMENT INDICATOR (EA18) VOLA	EKEUSESIG
EK INDUSTRY SURVEY: ORDER BOOK POSITION (EA) NADJ	EKEUSIOBR
EK SERVICES SVY- TOTAL: DEMAND - NEXT QTR (EA) SADJ	EKEUSNDXQ
EK INDUSTRIAL PRODUCTION: MANUFACTURING (EA18) VOLA	EKIPMAN.G
EK INDUSTRIAL PRODUCTION EXCLUDING CONSTRUCTION (EA18) VOLA	EKIPTOT.G
EK IND.: INTERMEDIATE GDS - EXP ORD BOOKS SADJ	EKITM3.BQ
EK IND.: INTERMEDIATE GDS - PROD EXPECT SADJ	EKITM5.BQ
EK IND.: INTERMEDIATE GDS - SELL PRICE EXPECT SADJ	EKITM6.BQ
EK IND.: INTERMEDIATE GDS - EMPL EXPECT SADJ	EKITM7.BQ
EK IND.: INVESTMENT GDS - EXP ORD BOOKS SADJ	EKIVE3.BQ
EK IND.: INVESTMENT GDS - PROD EXPECT SADJ	EKIVE5.BQ
EK IND.: INVESTMENT GDS - SELL PRICE EXPECT SADJ	EKIVE6.BQ
EK IND.: INVESTMENT GDS - EMPL EXPECT SADJ	EKIVE7.BQ
EK MFG - ORDER BOOKS: LEVEL SADJ	EKOBS078Q
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
EK RETAIL SURVEY: PRESENT BUSINESS SITUATION (EA) SADJ	EKTTR1BSQ
EK RETAIL SURVEY: VOLUME OF STOCKS (EA) SADJ	EKTTR2BSQ
EK RETAIL SURVEY: INTENTIONS OF PLACING ORDERS (EA) SADJ	EKTTR3BSQ
EK RETAIL SURVEY: EXPECTED BUSINESS SITUATION (EA) SADJ	EKTTR4BSQ
EK RETAIL SURVEY: EMPLOYMENT EXPECTATIONS (EA) SADJ	EKTTR5BSQ
EK RETAIL SURVEY: CONFIDENCE (EA) SADJ	EKTTR99BQ
EK BOP: CURRENT ACCOUNT - GOODS (NET) CURN	EKVISBOPA
EM EURO INTERBANK OFFERED RATE - 3-MONTH	EMIBOR3.
US CORP BONDS MOODYS SEASONED AAA (D) - MIDDLE RATE	FRCBAAA
US CORP BONDS MOODYS SEASONED BAA (D) - MIDDLE RATE	FRCBBAA
FTSE GLOBAL 100 ($) - PRICE INDEX	FTSEGL$
NYSE COMPOSITE - PRICE INDEX	NYSEALL
S&P 500 COMPOSITE - PRICE INDEX	S&PCOMP
SPREAD GERMANY: BD GOVERNMENT BOND YIELD 9-10 YRS MINUS BD 3-M	BDGBOND. -
FIBOR	BDOIR076R
UK MONTH AVERAGE SPOT EXCHANGE RATE, EUR INTO USD VOLN	UKAERD..P
UK LONDON GOLD PRICE - P.M. FIXING (EP)	UKGOLDP.
US CPI - ALL URBAN: ALL ITEMS SADJ	USCONPRCE
US FEDERAL FUNDS RATE (MONTHLY AVERAGE)	USFDFUND
US TREASURY BILL RATE - 3 MONTH (EP)	USGBILL3
US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA	USHOUSE.O
US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	USIPTOT.G
US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	USSHRPRCF

US STANDARD AND POORS' 500 COMPOSITE - DIVIDEND YLD	USSPDIVY
US UMICH CSS: CONSUMER SENTIMENT - EXPECTATIONS VOLN	USUMCONEH
US UNEMPLOYED (16 YRS & OVER) VOLA	USUNPTOTO
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WD COMMODITY PRICES: ALL COMMODITIES NADJ	WDI76ACDF
WD COMMODITY PRICES: ENERGY NADJ	WDI76ENDF
WILSHIRE 5000 TOTAL MARKET - PRICE INDEX	WIL5TMK
GERMANY	
BD NEW PASSENGER CAR REGISTRATIONS VOLN	BDCAR...P
BD CONSUMER CONFIDENCE INDICATOR - GERMANY SADJ	BDCNFCONQ
BD IND.: CONSUMER GDS - SELL PRICE EXPECT SADJ	BDCNS6.BQ
BD CONSTRUCTION ORDERS RECEIVED VOLN	BDCNSTR.H
BD CPI: TOTAL (FLASH & FINAL) NADJ	BDCONPRCF
BD NEW ORDERS TO MANUFACTURING - DOMESTIC: CONSUMER GOODS VOLA	BDDCNORDG
BD NEW ORDERS TO MANUFACTURING - DOMESTIC: CAPITAL GOODS VOLA	BDDCPORDG
BD EMPLOYMENT DURATION - SHORT-TERM WORKERS VOLN	BDEMPSTWP
BD CONSTRUCTION PRODUCTION INDEX: CONSTRUCTION VOLA	BDESCONMG
BD INDUSTRY T/O INDEX: MIG - CONSUMER GOODS SADJ	BDESCYEWE
BD CONSTRUCTION PRODUCTION INDEX: CIVIL ENGINEERING WORKS VOLA	BDESG1GZG
BD IPI: MANUFACTURE OF MOTOR VEHICLES VOLA	BDESI29MG
BD IPI: MIG - ENERGY (EXCEPT D AND E) VOLA	BDESIENXG
BD IPI: MIG - CONSUMER GOODS VOLA	BDESISURG
BD INDUSTRIAL PRODUCTION - INTERMEDIATE GOODS VOLA	BDESPIERG
BD INDUSTRIAL PRODUCTION - CAPITAL GOODS VOLA	BDESPIESG
BD INDUSTRY T/O INDEX: MIG - INTERMEDIATE GOODS SADJ	BDESRF78E
BD INDUSTRY T/O INDEX: MIG - CAPITAL GOODS SADJ	BDESSSOOE
BD UNEMPLOYMENT RATE, TOTAL SADJ	BDESUNEMO
BD TOTAL EXPORTS OF GOODS CURN	BDEXPBOPA
BD EXPORTS FOB (PAN BD M0790) CURN	BDEXPGDSA
BD LONG TERM GOVERNMENT BOND YIELD - 9-10 YEARS	BDGBOND.
BD CONSTRUCTION ORDERS RECEIVED - RESIDENTIAL CONSTRUCTION VOLA	BDHOUSE.G
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
BD CNSTR.IND.: CAPACITY UTILIZATION SADJ	BDIFDCTNQ
BD MFG. CONS. DURB.: BUSINESS SIT. NEXT 6 MO. - BAL. SADJ	BDIFDMDKQ
BD MFG. CONS. NON-DURB.: BUSINESS SIT. NEXT 6 MO. - BAL. SADJ	BDIFDMNKQ
BD MFG. CAPITAL GOODS: BUSINESS SIT. - BAL. SADJ	BDIFDMPAQ
BD MFG. CAPITAL GOODS: INVTRY. OF FINISHED GOODS - BAL. SADJ	BDIFDMPCQ
BD MFG. CAPITAL GOODS: BUSINESS SIT. NEXT 6 MO. - BAL. SADJ	BDIFDMPKQ
BD RET.SALE - DURABLES: ASSESSMENT OF BUSINESS SITUATION SADJ	BDIFRDUAQ
BD RET.SALE - DURABLES: ASSESSMENT OF INVENTORIES SADJ	BDIFRDUCQ
BD RET.SALE - NON-DURABLES: ASSESSMENT OF INVENTORIES SADJ	BDIFRNDCQ
BD RET.SALE-BUSINESS EXPECT.NEXT 6 MO.: INCL.MOT.VEH,PETR.ST.	BDIFRS.LR
BD RET.SALE - ASSESSMENT OF BUSINESS SITUATION SADJ	BDIFRSAAQ
BD WHOLESALE OF CONSUMER GOODS: ASSESSMENT OF BUSINESS SITUATION	BDIFWCOAQ
BD WHOLESALE OF FBT: EXPECT.OF BUSINESS DEVELOPMENT IN NEXT 6 MO	BDIFWFBLQ
BD TOTAL IMPORTS OF GOODS CURN	BDIMPBOPA
BD IMPORTS CIF (PAN BD M0790) CURN	BDIMPGDSA
BD INDUSTRIAL PRODUCTION - BASIC IRON & STEEL VOLN	BDIPIRSTH
BD INDUSTRIAL PRODUCTION INCLUDING CONSTRUCTION (CAL ADJ) VOLA	BDIPTOT.G
BD HWWI IDX OF WORLD MKT.PRC.OF RAW MATS,EURO AREA: EXCL.ENERGY	BDIUW501F
BD HWWI INDEX OF WORLD MKT.PRC.OF RAW MATS.FOR EURO AREA: ENERGY	BDIUW510F
BD MONEY SUPPLY-GERMAN CONTRIBUTION TO EURO M1(PAN BD M0790)	BDM1....A
BD MONEY SUPPLY-GERMAN CONTRIBUTION TO EURO M2(PAN BD M0690)	BDM2....A
BD MONEY SUPPLY-GERMAN CONTRIBUTION TO EURO M3(PAN BD M0690)	BDM3....A
BD YD 10-YR GOVT.BONDS(PROXY-9-10+ YEAR FEDERAL SECURITIES) NADJ	BDMIR080R
BD HARMONISED UNEMPLOYMENT: LEVEL, ALL PERSONS (ALL AGES) VOLA	BDMLFT15O
BD NEW ORDERS: MFG, MANUFACTURE OF ELECTRICAL EQUIP. , DOM. VOLA	BDNODEEQG
BD NEW ORDERS: MFG, MOTOR VEH., TRAILERS&SEMI-TRAIL., DOM. SADJ	BDNODVEME
BD NEW ORDERS: MFG,MOTOR VEH., TRAILERS&SEMI-TRAIL., ABROAD SADJ	BDNOFVEME
BD REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	BDOCC011
BD NEW ORDERS TO MFG. - FROM ABROAD: CONSUMER GOODS VOLA	BDOCNORDG
BD DEU CPI ENERGY NADJ	BDOCP041F
BD NEW ORDERS TO MANUFACTURING - FROM ABROAD: CAPITAL GOODS VOLA	BDOCPORDG
BD 3-MONTH FIBOR NADJ	BDOIR076R
BD CLI SPREAD OF IRS NADJ	BDOL2056R
BD ORDERS FOR MFC INTERMEDIATE GOODS (VOLUME) VOLA	BDOODI51G
BD SHARE PRICES: CDAX INDEX NADJ	BDOSP001F
BD PPI: MOTOR FUEL, INCL. AIRCRAFT FUEL NADJ	BDPPFUMTF
BD DISCOUNT RATE / SHORT TERM EURO REPO RATE	BDPRATE.
BD PPI: INDL. PRODUCTS, TOTAL, SOLD ON THE DOMESTIC MARKET NADJ	BDPROPRCF
BD DAX SHARE PRICE INDEX, EP NADJ	BDSHRPRCF
BD IND.T/O: COMPUTER, ELECTRONIC & OPTICAL PRODUCTS, DOM. VOLA	BDSTDCEOG
BD IND.T/O: MOTOR VEH., TRAILERS&SEMI-TRAIL., DOM. VOLA	BDSTDVEMG
BD IND. T/O: CAPITAL GOODS, FGN. SADJ	BDSTFCAPE
BD IND.T/O: COMPUTER, ELECTRONIC & OPTICAL PRODUCTS, FGN. VOLN	BDSTFCEOH
BD IND. T/O: CONS. GOODS, FGN. SADJ	BDSTFCONE

BD IND. T/O: INTERMEDIATE GOODS, FGN. SADJ	BDSTFINTE
BD IND.T/O: MOTOR VEH., TRAILERS&SEMI-TRAIL., FGN. VOLN	BDSTFVEMH
BD MFG ORDERS: COMPUTER,ELECC.&OPT.PRDS,ELECL. EQUIP., FGN. VOLA	BDUSC588G
BD CONSTRUCTION ORDERS RECEIVED-NON-RESIDENTIAL CONSTRUCTION	BDUSDA19G
BD CONSTRUCTION ORDERS RECEIVED - CIVIL ENGINEERING VOLA	BDUSDA20G
BD CPI (EXCLUDING ENERGY) SADJ	BDUSFB76E
BD CONSTRUCTION: MAN-HOURS WORKED CURA	BDUSMB11B
BD TURNOVER IN CONSTRUCTION - TOTAL CURA	BDUSMB28B
BD TURNOVER IN CONSTRUCTION - INDUSTRIAL BUILDING CURA	BDUSMB31B
BD TURNOVER IN CONSTRUCTION - RESIDENTIAL BUILDING CURA	BDUSMB36B
BD TURNOVER IN CONSTRUCTION- PUB SECTOR & ROAD CONSTRUCTION CURA	BDUSMB37B
BD VACANCIES VOLN	BDUUCC04P
BD REX PRICE INDEX, END OF MONTH CURN	BDWU035AA
BD FULLY TAXED BONDS OUTSTANDING-YIELDS:GT1Y&UP TO 2Y TO MATURIT	BDWU0898R
BD FULLY TAXED BONDS OUTSTANDING-YIELDS:GT5Y&UP TO 6Y TO MATURIT	BDWU0902R
BD FULLY TAXED BONDS OUTSTAND-YIELDS GT 9Y& UP TO 10Y TO MTRY.	BDWU8608R
BD YLDS ON LSTED FEDRL BNDS OUTSTNDG.MATURITY 3-5 YRS AVE.RATE	BDWU9552R
BD YIELDS ON LISTED FED.SEC,DER.FM.TS OF IR,RESID.MTRY.1 YR.	BDWZ3400
CH EXPORTS CURN	CHEXPGDSA
CH INDUSTRIAL PRODUCTION INDEX VOLN	CHIPTOT.H
CH GOLD AND FOREIGN RESERVES - FOREIGN RESERVE CURN	CHRESERVA
EURO STOXX - PRICE INDEX	DJEURST
EK CONSUMER CONFIDENCE INDICATOR (EA) SADJ	EKCNFCONQ
EK INDUSTRIAL PRODUCTION - ENERGY (EA18) VOLA	EKESIENGG
EK INDUSTRY SURVEY: PRODUCTION EXPECTATIONS (EA) SADJ	EKEUIMF5Q
EK INDUSTRIAL PRODUCTION: MANUFACTURING (EA18) VOLA	EKIPMAN.G
EK INDUSTRIAL PRODUCTION EXCLUDING CONSTRUCTION (EA18) VOLA	EKIPTOT.G
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
US CORP BONDS MOODYS SEASONED AAA (D) - MIDDLE RATE	FRCBAAA
US CORP BONDS MOODYS SEASONED BAA (D) - MIDDLE RATE	FRCBBAA
NYSE COMPOSITE - PRICE INDEX	NYSEALL
S&P 500 COMPOSITE - PRICE INDEX	S&PCOMP
SPREAD GERMANY: BD GOVERNMENT BOND YIELD 9-10 YRS MINUS BD 3-M FIBOR	BDGBOND. - BDOIR076R
UK MONTH AVERAGE SPOT EXCHANGE RATE, EUR INTO USD VOLN	UKAERD..P
UK LONDON GOLD PRICE - P.M. FIXING (EP)	UKGOLDP.
US CPI - ALL URBAN: ALL ITEMS SADJ	USCONPRCE
US FEDERAL FUNDS RATE (MONTHLY AVERAGE)	USFDFUND
US TREASURY BILL RATE - 3 MONTH (EP)	USGBILL3
US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA	USHOUSE.O
US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	USIPTOT.G
US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	USSHRPRCF
US STANDARD AND POORS' 500 COMPOSITE - DIVIDEND YLD	USSPDIVY
US UMICH CSS: CONSUMER SENTIMENT - EXPECTATIONS VOLN	USUMCONEH
US UNEMPLOYED (16 YRS & OVER) VOLA	USUNPTOTO
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WILSHIRE 5000 TOTAL MARKET - PRICE INDEX	WIL5TMK
FRANCE	
BD LONG TERM GOVERNMENT BOND YIELD - 9-10 YEARS	BDGBOND.
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
BD DISCOUNT RATE / SHORT TERM EURO REPO RATE	BDPRATE.
CH GOLD AND FOREIGN RESERVES - FOREIGN RESERVE CURN	CHRESERVA
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
FR CAPITAL MARKET YIELDS-BOND YIELD,PRIVATE SECTOR(AVERAGE) NADJ	FR101066
FR WHOLESALE COMPOSITE INDICATOR NADJ	FR579137F
FR SURVEY - HOUSEHOLDS, ECONOMIC SITUATION PAST 12M SADJ	FR857188Q
FR SURVEY - HOUSEHOLDS, ECONOMIC SITUATION NEXT 12M SADJ	FR857189Q
FR SURVEY-HOUSEHOLD OPINION ON FUTURE EVOLUTION ON UNEMPLMT.	FR857190Q
FR SURVEY-HOUSEHOLD OPINION ON PAST EVOLUTION OF CONSUMER PRICES	FR857191Q
FR SURVEY-HOUSEHOLD OPINION ON FUTURE EVOLUTION CONSUMER PRICES	FR857192Q
FR SURVEY-HOUSEHOLD OPINION ON IMPORTANT PURCHASE INTENTIONS	FR857193Q
FR SURVEY - HOUSEHOLD OPINION ON PAST FINANCIAL SITUATION SADJ	FR857196Q
FR SURVEY - HOUSEHOLD OPINION ON FUTURE FINANCIAL SITUATION SADJ	FR857197Q
FR MFI LOANS TO RESIDENT PRIVATE SECTOR CURN	FRBANKLPA
FR BDF ASSETS: FRANCE - LOANS CENTRAL GOVERNMENT CURN	FRBDFFCGA
US CORP BONDS MOODYS SEASONED AAA (D) - MIDDLE RATE	FRCBAAA
FR SURVEY: MANUFACTURING OUTPUT LEVEL - GENERAL OUTLOOK SADJ	FRCNFBUSQ
FR UNEMPLOYMENT (HARMONIZED): TOTAL TRND	FRESQT8JT
FR UNEMPLOYMENT RATE, TOTAL SADJ	FRESUNEMO
FR CAPITAL MARKET YIELDS-13-WEEK TREASURY BILLS,MO.WGHTD.AVG.	FRGBILL3
FR GOVERNMENT GUARANTEED BOND YIELD (EP) NADJ	FRGBOND.
FR HOUSEHOLD CONSUMPTION - AUTOMOBILES CONA	FRHCONAUD
FR HOUSEHOLD CONSUMPTION - DURABLE GOODS CONA	FRHCONDGD
FR HOUSEHOLD CONSUMPTION - HOUSEHOLD EQPT CONA	FRHCONLGD

FR HOUSEHOLD CONSUMPTION - MANUFACTURED GOODS CONA	FRHCONMFD
FR HOUSEHOLD CONSUMPTION - ENGINEERED PRODUCTS CONA	FRHCONMGD
FR HOUSEHOLD CONSUMPTION - TEXTILES & LEATHER CONA	FRHCONTLD
FR INDUSTRIAL PRODUCTION VOLA	FRIPTOT.G
FR MFI LIABILITIES: FRANCE - DEPOSITS CENTRAL GOVERNMENT CURN	FRMFIDFGA
FR MFI ASSETS: LOANS - GENERAL GOVERNMENT CURN	FRMFIFCGA
FR REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	FROCC011
FR 3-MONTH PIBOR NADJ	FROIR076R
FR AVERAGE COST OF FUNDS FOR BANKS / EURO REPO RATE	FRPRATE.
FR SHARE PRICE INDEX - SBF 250 NADJ	FRSHRPRCF
FR BANQUE DE FRANCE SVY.: BUSINESS SENTIMENT INDICATOR(CAL ADJ)	FRSURCBSQ
FR SURVEY: MFG. OUTPUT - ORDER BOOK & FOREIGN DEMAND SADJ	FRSURFMPQ
FR SURVEY: MANUFACTURING OUTPUT - ORDER BOOK & DEMAND SADJ	FRSURGMPQ
FR SURVEY: MFG. OUTPUT - FINISHED GOODS INVENTORIES SADJ	FRSURSMPQ
FR US $ TO 1 EURO (FRENCH FRANC DERIVED HISTORY PRIOR 1999)	FRXRUSE.
NYSE COMPOSITE - PRICE INDEX	NYSEALL
S&P 500 COMPOSITE - PRICE INDEX	S&PCOMP
SPREAD: FR GOVERNMENT BOND YIELD MINUS FR CAPITAL MKT YIELDS-13-WEEK T BILLS	FRGBOND.- FRGBILL3
UK MONTH AVERAGE SPOT EXCHANGE RATE, EUR INTO USD VOLN	UKAERD..P
UK LONDON GOLD PRICE - P.M. FIXING (EP)	UKGOLDP.
US CPI - ALL URBAN: ALL ITEMS SADJ	USCONPRCE
US FEDERAL FUNDS RATE (MONTHLY AVERAGE)	USFDFUND
US TREASURY BILL RATE - 3 MONTH (EP)	USGBILL3
US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA	USHOUSE.O
US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	USIPTOT.G
US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	USSHRPRCF
US STANDARD AND POORS' 500 COMPOSITE - DIVIDEND YLD	USSPDIVY
US UMICH CSS: CONSUMER SENTIMENT - EXPECTATIONS VOLN	USUMCONEH
US UNEMPLOYED (16 YRS & OVER) VOLA	USUNPTOTO
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WILSHIRE 5000 TOTAL MARKET - PRICE INDEX	WIL5TMK
ITALY	
ITALY-DS Automobiles - PRICE INDEX	AUTOSIT
Baltic Exchange Dry Index (BDI) - PRICE INDEX	BALTICF
ITALY-DS Banks - PRICE INDEX	BANKSIT
GERMANY BENCHMARK BOND 10 YR (DS) - RED. YIELD	BDBRYLD
BD LONG TERM GOVERNMENT BOND YIELD - 9-10 YEARS	BDGBOND.
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
MNY MKT - 3-MONTH FRANKFURT 'DEAD' - MIDDLE RATE	BDMNY3M
CH EXPORTS CURN	CHEXPGDSA
CH INDUSTRIAL PRODUCTION INDEX VOLN	CHIPTOT.H
CH GOLD AND FOREIGN RESERVES - FOREIGN RESERVE CURN	CHRESERVA
DC WORLD BNK NON-ENERGY COMMODITIY PRICES: LOWER MID INCM CNTRYS	DCI76AXDF
EURO STOXX - PRICE INDEX	DJEURST
EK CONSUMER CONFIDENCE INDICATOR (EA) SADJ	EKCNFCONQ
EK INDUSTRY SURVEY: PRODUCTION EXPECTATIONS (EA) SADJ	EKEUIMF5Q
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
ITALY-DS Electricity - PRICE INDEX	ELECTIT
ITALY-DS Eltro/Elec Eq - PRICE INDEX	ELTNCIT
US CORP BONDS MOODYS SEASONED AAA (D) - MIDDLE RATE	FRCBAAA
US CORP BONDS MOODYS SEASONED BAA (D) - MIDDLE RATE	FRCBBAA
ITALY-DS Industrials - PRICE INDEX	INDUSIT
ITALY-DS Tch H/W & Eq - PRICE INDEX	INFOHIT
ITALY-DS Insurance - PRICE INDEX	INSURIT
IT BUSINESS SVY.: ECONOMY IN NEXT 3MOS., NET NADJ	ITBSINEVR
IT BUSINESS SVY.: SELLING PRICE IN NEXT 3MOS., NET NADJ	ITBSINSBR
ITALY T-BILL AUCT. GROSS 3 MONTH - MIDDLE RATE	ITBT03G
ITALY T-BILL AUCT. GROSS 6 MONTH - MIDDLE RATE	ITBT06G
ITALY T-BILL AUCT. GROSS 12 MONTH - MIDDLE RATE	ITBT12G
IT NEW PASSENGER CAR REGISTRATIONS VOLN	ITCAR...P
IT BUS.SVY.: CONSUMER GOODS - ORDER BOOKS EXPORT, NET NADJ	ITCONEXOR
IT BUS.SVY.: CONS.GDS.- ECONOMY IN NEXT 3MOS., NET NADJ	ITCONFECR
IT BUS.SVY.: CONS.GDS.- ORDER BOOKS IN NEXT 3MOS., NET NADJ	ITCONFOBR
IT BUS.SVY.: CONS.GDS.- PRODUCTION IN NEXT 3MOS., NET NADJ	ITCONFPRR
IT BUS.SVY.: CONS.GDS.- SELLING PRICE IN NEXT 3MOS., NET NADJ	ITCONFSPR
IT BUS.SVY.: CONSUMER GOODS - ORDER BOOKS DOMESTIC, NET NADJ	ITCONOBBR
IT BUS.SVY.: CONSUMER GOODS - ORDER BOOKS, NET NADJ	ITCONORDR
IT BUS.SVY.: CONSUMER GOODS - PRODUCTION LEVEL, NET NADJ	ITCONPRDR
IT BUS.SVY.: CONSUMER GOODS - STOCKS OF FIN.GDS., NET NADJ	ITCONSFGR
IT BUSINESS SVY.: ORDER BOOKS - DOMESTIC, NET NADJ	ITDOMORDR
IT NEW RESIDENTIAL BUILDINGS - COST INDEX NADJ	ITECEIBCF
IT LONG TERM GOVT.BOND YIELDS- MAASTRICHT DEFINITION (AVG.)	ITESEFIGR
IT ENERGY - CONSUMPTION, ELECTRICITY VOLN	ITESEIWAP
IT UNEMPLOYMENT VOLA	ITESTUNPO
IT UNEMPLOYMENT RATE, TOTAL SADJ	ITESUNEMO
IT ECONOMIC SENTIMENT INDICATOR VOLA	ITEUSESIG

IT BUSINESS SVY.: ORDER BOOKS IN NEXT 3MOS., NET NADJ	ITEXPORDR
IT BUSINESS SVY.: ORDER BOOKS - EXPORT, NET NADJ	ITFORORDR
IT GOVERNMENT BOND GROSS YIELD (RENDISTATO) (EP)	ITGBOND.
IT BUS.SVY.: INTERMED. GDS.- ORDER BOOKS EXPORT, NET NADJ	ITINTEXOR
IT BUS.SVY.: INTERMED.GDS.- ECONOMY IN NEXT 3MOS., NET NADJ	ITINTFECR
IT BUS.SVY.: INTERMED.GDS.- ORDER BOOKS IN NEXT 3MOS., NET NADJ	ITINTFOBR
IT BUS.SVY.: INTERMED.GDS.- PRODUCTION IN NEXT 3MOS., NET NADJ	ITINTFPRR
IT BUS.SVY.: INTERMED.GDS.- SELL.PRICE IN NEXT 3MOS., NET NADJ	ITINTFSPR
IT BUS.SVY.: INTERMED. GDS.- ORDER BOOKS DOMESTIC, NET NADJ	ITINTOBBR
IT BUS.SVY.: INTERMED. GDS.- ORDER BOOKS, NET NADJ	ITINTORDR
IT BUS.SVY.: INTERMED. GDS.- PRODUCTION LEVEL, NET NADJ	ITINTPRDR
IT BUS.SVY.: INTERMED. GDS.- STOCKS OF FIN.GDS., NET NADJ	ITINTSFGR
IT BUS.SVY.:INVESTMENT GOODS - ORDER BOOKS EXPORT, NET NADJ	ITINVEXOR
IT BUS.SVY.: INV.GDS.- ECONOMY IN NEXT 3MOS., NET NADJ	ITINVFECR
IT BUS.SVY.: INV.GDS.- ORDER BOOKS IN NEXT 3MOS., NET NADJ	ITINVFOBR
IT BUS.SVY.: INV.GDS.- PRODUCTION IN NEXT 3MOS., NET NADJ	ITINVFPRR
IT BUS.SVY.: INV.GDS.- SELLING PRICE IN NEXT 3MOS., NET NADJ	ITINVFSPR
IT BUS.SVY.:INVESTMENT GOODS - ORDER BOOKS DOMESTIC, NET NADJ	ITINVOBBR
IT BUS.SVY.:INVESTMENT GOODS - ORDER BOOKS, NET NADJ	ITINVORDR
IT BUS.SVY.:INVESTMENT GOODS - PRODUCTION LEVEL, NET NADJ	ITINVPRDR
IT BUS.SVY.:INVESTMENT GOODS - STOCKS OF FIN.GDS., NET NADJ	ITINVSFGR
IT INDUSTRIAL PRODN.: ELECTRICITY,GAS,STEAM & AIR CONDITIONED	ITIP350EG
IT INDUSTRIAL PRODUCTION: CONSUMER GOODS - DURABLE VOLA	ITIPCGDRG
IT INDUSTRIAL PRODUCTION: CONSUMER GOODS - NON-DURABLE VOLA	ITIPCGNDG
IT INDUSTRIAL PRODN.: CHEMICAL PRODUCTS & SYNTHETIC FIBRES VOLA	ITIPCHEMG
IT INDUSTRIAL PRODN.: COMPUTER,ELECTRONIC & OPTICAL PRODS. VOLA	ITIPCI0EG
IT INDUSTRIAL PRODUCTION: ELECTRICAL EQUIPMENT VOLA	ITIPCJ0QG
IT INDUSTRIAL PRODN.: OTH.MFG,& REPAIR & INSTALL-MACH & EQP VOLA	ITIPCMOMG
IT INDUSTRIAL PRODUCTION: CONSUMER GOODS VOLA	ITIPCNGDG
IT INDUSTRIAL PRODUCTION: ENERGY VOLA	ITIPENGYG
IT INDUSTRIAL PRODUCTION: FOOD, DRINK & TOBACCO VOLA	ITIPFOODG
IT INDUSTRIAL PRODN.: COKE MANUFACTURE & PETROLEUM REFINING VOLA	ITIPFUELG
IT INDUSTRIAL PRODUCTION: INTERMEDIATE GOODS VOLA	ITIPINTMG
IT INDUSTRIAL PRODUCTION: INVESTMENT GOODS VOLA	ITIPINVTG
IT INDUSTRIAL PRODUCTION: MACHINES & MECHANICAL APPARATUS VOLA	ITIPMACHG
IT INDUSTRIAL PRODUCTION: MANUFACTURING VOLA	ITIPMAN.G
IT INDUSTRIAL PRODUCTION: METAL & METAL PRODUCTS VOLA	ITIPMETLG
IT INDUSTRIAL PRODUCTION: EXTRACTION OF MINERALS VOLA	ITIPMINGG
IT INDUSTRIAL PRODUCTION: BASIC PHARMACEUTICAL PRODUCTS VOLA	ITIPPHARG
IT INDUSTRIAL PRODUCTION: RUBBER ITEMS & PLASTIC MATERIALS VOLA	ITIPRUBRG
IT INDUSTRIAL PRODUCTION: TEXTILE & CLOTHING VOLA	ITIPTEXTG
IT INDUSTRIAL PRODUCTION VOLA	ITIPTOT.G
IT INDUSTRIAL PRODUCTION: MEANS OF TRANSPORT VOLA	ITIPTRNSG
IT INDUSTRIAL PRODUCTION: WOOD & WOOD PRODUCTS VOLA	ITIPWOODG
IT BUSINESS SVY.: STOCKS OF FINISHED GOODS, NET NADJ	ITLEVINVR
IT MONEY SUPPLY: M1 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	ITM1....A
IT MONEY SUPPLY: M2 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	ITM2....A
IT MONEY SUPPLY: M3 - ITALIAN CONTRIBUTION TO THE EURO AREA CURN	ITM3....A
IT CAR REGISTRATIONS-NEW MEDIUM & HEAVY CML.VEHICLES OVER 3.5T	ITMCVREGP
FTSE ITALIA MIB STORICO - PRICE INDEX	ITMHIST
IT NEW ORDERS SADJ	ITNEWORDE
IT REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	ITOCC011
IT 3-MONTH INTERBANK RATE ON DEPOSITS NADJ	ITOIR076R
IT DISCOUNT RATE / SHORT TERM EURO REPO RATE	ITPRATE.
IT BUSINESS SVY.: PRODUCTION IN NEXT 3MOS., NET NADJ	ITPRDEXPR
IT PPI - LINKED & REBASED NADJ	ITPROPRAF
IT MILAN COMIT GENERAL SHARE PRICE INDEX (EP) NADJ	ITSHRPRCF
IT BUSINESS SVY.: ORDER BOOKS, NET NADJ	ITTOTORDR
IT BUSINESS SVY.: PRODUCTION LEVEL, NET NADJ	ITTOTPRDR
IT IND.: OVERALL - EMPL EXPECT SADJ	ITTTA7BSQ
IT ITALIAN LIRE TO US $ (MTH.AVG.)	ITXRUSD.
ITALY-DS Media - PRICE INDEX	MEDIAIT
NYSE COMPOSITE - PRICE INDEX	NYSEALL
ITALY-DS Oil & Gas - PRICE INDEX	OILGSIT
ITALY-DS Pharm & Bio - PRICE INDEX	PHARMIT
S&P 500 COMPOSITE - PRICE INDEX	S&PCOMP
SPREAD: IT GOVERNMENT BOND GROSS YIELD MINUS ITALY T-BILL 3 M RATE	ITGBOND.-ITBT03G
ITALY-DS Telecom - PRICE INDEX	TELCMIT
ITALY-DS Market - PRICE INDEX	TOTMKIT
UK MONTH AVERAGE SPOT EXCHANGE RATE, EUR INTO USD VOLN	UKAERD..P
UK LONDON GOLD PRICE - P.M. FIXING (EP)	UKGOLDP.
US CPI - ALL URBAN: ALL ITEMS SADJ	USCONPRCE
US FEDERAL FUNDS RATE (MONTHLY AVERAGE)	USFDFUND
US TREASURY BILL RATE - 3 MONTH (EP)	USGBILL3
US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA	USHOUSE.O
US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	USIPTOT.G
US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	USSHRPRCF
US STANDARD AND POORS' 500 COMPOSITE - DIVIDEND YLD	USSPDIVY
US UMICH CSS: CONSUMER SENTIMENT - EXPECTATIONS VOLN	USUMCONEH

US UNEMPLOYED (16 YRS & OVER) VOLA	USUNPTOTO
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WD COMMODITY PRICES: FOOD NADJ	WDI76EXDF
WD COMMODITY PRICES: NON-ENERGY NADJ	WDI76NFDF
WILSHIRE 5000 TOTAL MARKET - PRICE INDEX	WIL5TMK
UK	
BD LONG TERM GOVERNMENT BOND YIELD - 9-10 YEARS	BDGBOND.
BD HWWA INDEX OF WORLD MARKET PRICES OF RAW MATS, EURO AREA NADJ	BDHWWAINF
EK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	EKOCC011
NYSE COMPOSITE - PRICE INDEX	NYSEALL
S&P 500 COMPOSITE - PRICE INDEX	S&PCOMP
UK YIELD 10-YEAR MINUS UK DISCOUNT RATE 3-MONTH TREASURY BILLS	UKOIR080R - UKOIR077R
UK MONTH AVERAGE SPOT EXCHANGE RATE, EUR INTO USD VOLN	UKAERD..P
UK TREASURY BILLS: DISCOUNT RATE 3M	UKAJNB..
UK MONTHLY AVERAGE SPOT EXCHANGE RATES, STERLING INTO EURO	UKASERM.
UK AVERAGE SPOT EXCHANGE RATE: UK TO US $ VOLN	UKAUSS..P
UK TOTAL CLAIMANT COUNT VOLA	UKBCJD..
UK CLAIMANT COUNT RATE, ALL SADJ	UKBCJE..
UK CBI MONTHLY ENQUIRY: VOLUME OF EXPECTED OUTPUT - BALANCE NADJ	UKCBIOPB
UK LSL/ ACAD AVERAGE HOUSE PRICE CURA	UKFTHPI.B
UK GROSS REDEMPTION YIELD ON 20 YEAR GILTS (PERIOD AVERAGE) NADJ	UKGBOND.
UK GFK CONSUMER CONFIDENCE INDEX NADJ	UKGFKCCNR
UK LONDON GOLD PRICE - P.M. FIXING (EP)	UKGOLDP.
UK INTERBANK RATE - 3 MONTH (MONTH AVG)	UKINTER3
UK INDUSTRIAL PRODUCTION INDEX - MANUFACTURING VOLA	UKIPMAN.G
UK INDEX OF PRODUCTION - ALL PRODUCTION INDUSTRIES VOLA	UKIPTOT.G
UK IOP: B: MINING AND QUARRYING VOLA	UKK224..G
UK IOP: CB: MFG.OF TEXTILES,WEARING APPAREL & LEATHER PRODS.	UKK22P..G
UK IOP: CC: MANUFACTURE OF WOOD & PAPER PRODS, AND PRINTING VOLA	UKK22T..G
UK IOP: CE: MANUFACTURE OF CHEMICALS AND CHEMICAL PRODUCTS VOLA	UKK22Z..G
UK IOP: CG: MFG.RUBBER & PLASTICS PRODS,& OTHER NON-METALLIC	UKK23B..G
UK IOP: CH: MANUFACTURE OF BASIC METALS AND METAL PRODUCTS VOLA	UKK23G..G
UK IOP: CI:MANUFACTURING OF COMPUTER,ELECTRONIC & OPTICAL PRODS.	UKK23N..G
UK IOP: CK: MANUFACTURE OF MACHINERY AND EQUIPMENT N.E.C. VOLA	UKK23R..G
UK IOP: CL: MANUFACTURE OF TRANSPORT EQUIPMENT VOLA	UKK23T..G
UK MONEY SUPPLY M0: NOTES & COINS IN CIRC.OUTSIDE BANK OF ENGLAN	UKM0....B
UK LFS: IN EMP.: AGED 16+: ANNUAL = SPRING QUARTER(MAR-MAY) VOLA	UKMGRZ..O
UK LFS: UNEMPLOYED: AGED 16+ VOLA	UKMGSC..O
UK REAL EFFECTIVE EXCHANGE RATES - CPI BASED VOLN	UKOCC011
UK EXCHANGE RATE MONTHLY AVERAGE NADJ	UKOCC015
UK DISCOUNT RATE 3-MONTH TREASURY BILLS (STERLING) NADJ	UKOIR077R
UK BANK OF ENGLAND BASE RATE (EP)	UKPRATE.
UK PPI - OUTPUT OF MANUFACTURED PRODUCTS (HOME SALES) NADJ	UKPROPRCF
UK RPI - ALL ITEMS EXCLUDING MORTGAGE INTEREST NADJ	UKRPAXMIF
UK FT ALL SHARE INDEX (EP) NADJ	UKSHRPRCF
UK UNEMPLOYMENT RATE SADJ	UKUN%TOTQ
UK JAPANESE YEN TO UK	UKXYEN..
UK LFS: TOTAL ACTUAL WEEKLY HOURS WORKED, ALL VOLA	UKYBUS..O
US CPI - ALL URBAN: ALL ITEMS SADJ	USCONPRCE
US EMPLOYED - NONFARM INDUSTRIES TOTAL (PAYROLL SURVEY) VOLA	USEMPALLO
US FEDERAL FUNDS RATE (MONTHLY AVERAGE)	USFDFUND
US TREASURY BILL RATE - 3 MONTH (EP)	USGBILL3
US NEW PRIVATE HOUSING UNITS STARTED (AR) VOLA	USHOUSE.O
zUS INDL PROD - AUTOMOBILE VOLA	USIP336AG
US GROSS VALUE OF PRODUCTION - AUTOMOTIVE PRODUCTS (AR) CONA	USIPAUTOD
US INDL PROD - DURABLE CONSUMER GOODS VOLA	USIPMDUCG
US INDL PROD - MFG (NAICS) VOLN	USIPMFG.H
US INDL PROD - NONDURABLE CONSUMER GOODS VOLA	USIPMNOCG
US INDUSTRIAL PRODUCTION - TOTAL INDEX VOLA	USIPTOT.G
US DOW JONES INDUSTRIALS SHARE PRICE INDEX (EP) NADJ	USSHRPRCF
US STANDARD AND POORS' 500 COMPOSITE - DIVIDEND YLD	USSPDIVY
US UNEMPLOYED (16 YRS & OVER) VOLA	USUNPTOTO
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WD COMMODITY PRICES: CRUDE OIL NADJ	WDI76AADF
WILSHIRE 5000 TOTAL MARKET - PRICE INDEX	WIL5TMK

Bibliography

Bai, J. (2003). Inferential theory for factor models of large dimensions, Econometrica 71: 135-173.

Bailey, N., Kapetanios, G. and Pesaran, M. H. (2013). Exponent of cross-sectional dependence: Estimation and inference, Mimeo, University of Cambridge .

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

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis?, 132: 169-194.

Capistrán, C. and Timmermann, A. (2008). Forecast combination with entry and exit of experts, Discussion Paper, 2008, available online at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.5652&rep=rep1&type=pdf.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis in large asset markets, 51: 1305-1324.

De Mol, C., Giannone, D. and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian regression a valid alternative to principal components?, 146: 318-328.

Elliott, G. and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions, 122: 47-79.

Faust, J. and Wright, J. H. (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset, mimeo, Federal Reserve Board and Johns Hopkins University.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation, 82: 540-554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). The generalised factor model: Consistency and rates, Journal of Econometrics 119: 231-255.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting, Journal of the American Statistical Association 100(471): 830-840. 15

Garthwaite, P. H. (1994). An interpretation of partial least squares, 89: 122-127.

Granger, C. W. J. and Ramanathan, R. (1984). Improved methods of combining forecasts, Journal of Forecasting 3: 197-204.

Groen, J. J. J. and Kapetanios, G. (2015), Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting, Federal Reserve Bank of New York, Staff Report No. 327, May 2008, Revised October 2015, available online at: https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr327.pdf.

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, Scandinavian Journal of Statistics 17: 97-114.

Hendry, D. F. (1995). Dynamic Econometrics, Oxford University Press.

Kapetanios, G. and Marcellino, M. (2010). Factor-gmm estimation with large sets of possibly weak instruments, Computational Statistics and Data Analysis 54(11): 2655-2675.

Kelly, B. and Pruitt, S. (2012). The three-pass regression filter: A new approach to forecasting using many predictors, Mimeo, University of Chicago.

Krolzig, H.-M. and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures, 25: 831-866.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, Philosophical Magazine 2: 559-572.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors, 97: 1167-1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes, 20: 147-162.

Svensson, L. E. O. (2005). Monetary policy with judgment: Forecast targeting, International Journal of Central Banking 1(1): 1-54.

Timmermann, A. (2006). Forecast combinations, in G. Elliott, C. W. J. Granger and A. Timmermann (eds), Handbook of Economic Forecasting, North Holland, Amsterdam, pp. 135-196. 16

Wold, H. (1982). Soft modeling. the basic design and some extensions, in K.-G. Jöreskog and H. Wold (eds), Systems Under Indirect Observation, Vol. 2, North-Holland, Amsterdam.

V Rapid Estimates based on Combining Forecasting Techniques

12 Combined Forecasting Methods and Rapid Estimates

Authors[†]

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

Table of Contents

[†]Note: The first version of this chapter was submitted in January 2013.

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

12.1 Introduction

The availability of multiple forecasts for the same variable reflects heterogeneity in the information sets and differences in forecasters' modeling strategies and subjective judgments, due to private information or dissimilarities in the assumptions underlying the forecasting exercise. Hence, multiple forecasts provide the opportunity to exploit a much richer information set than what is conventionally embedded in single time series forecasts.

The first issue that arises when multiple forecasts are available is how best to exploit all the existing information, and the idea of diversification gains motivates the proposal of combining the alternative forecasts. Generally speaking, the forecast user's information set is not the union of the information sets underlying the alternative available forecasts, but a much smaller subset. The less similar the subset to the union, the more useful a combination of forecasts will be.

Another reason that justifies forecast pooling, and the good performance of simple weighting schemes, is that using averages of many predictors may provide some robustness against structural instability (see, inter alia, Diebold and Pauly (1990), Clements and Hendry (1999), and Aiolfi and Timmermann (2004)). Since it is difficult to detect the presence of structural breaks in real time, on average combinations of forecasts from models with different degrees of adaptability to these breaks will outperform forecasts from individual models. In particular, most standard forecasts will fail in the same direction when forecasting over a period within which a break unexpectedly occurs, and the extent of the forecast failure depends on the degree of model misspecification, data correlation, and size and timing of the break. However, Clements and Hendry (1999) have shown that in this context combining alternative forecasts can act like an intercept correction, which puts the combined forecast back on track. Combining forecasts across different models can be also seen as an attempt to make forecasts more robust against misspecification biases of unknown form. In particular, if two forecasts are differently biased (e.g. upwards and downwards), their combination can be an improvement over either single forecast. This point has been particularly stressed by Diebold and Lopez (1996) and Stock and Watson (1999). Finally, it is possible to obtain different forecasts based on the same information set, but on different loss functions. As in Christoffersen and Diebold (1997), it is then possible to obtain a combination of the forecasts that outperforms the individual ones.

Of course, there are also arguments against using forecast pooling, in particular problems related to the estimation and the stability of the combination weights (e.g., see Diebold and Pauly (1990) for estimation problems when the sample size is small relative to the number of forecasts). Moreover, when the full information set is available, the use of a combination of forecasts based on different subsets of the available information can be suboptimal with respect to the use of a single model that exploits the whole information set. However, the specification and estimation of the latter model is typically very complicated, particularly in real time, see e.g. Diebold and Pauly (1990). Hence, whether forecast combination is useful or not for a specific application becomes mostly an empirical issue. From an empirical point of view, in their pioneering study, Bates and Granger (1969) provide an example of the usefulness of combining forecasts from linear and exponential trend models of output. Although the exponential model forecasts have a much smaller sum of squared errors than the linear model, a combination which attaches a small weight to the linear forecasts has a smaller sum of squared errors. After Bates and Granger (1969), several other studies have found that a combination of forecasts for economic variables may perform better than each one of the single constituents' forecast, in the sense that the combined forecasts delivers a smaller mean-squared forecasts error, see e.g. the reviews in Clemen (1989) and Timmermann (2006). Moreover, simple rules for combining forecasts, such as using equal weighs, often work as well as (or better than) more complex rules based on the relative past performance of the forecasts to be combined, see e.g. Stock and Watson (2004). The first section of this chapter provides a more detailed overview of the theoretical and empirical literature on forecast combination. In the second section, we undertake a careful evaluation of the empirical performance of a set of alternative forecast combination methods for the PEEIs.[1]

[1] Principal European Economic Indicators

12.2 An overview of forecast combination

We start this first section with a more formal definition of the problem of forecast combination (Section 12.2.1). Then we discuss linear combination (Section 12.2.2) and issues related to weight estimation (Section 12.2.3), techniques for the large set of forecasts case (Section 12.2.4), time-varying and non-linear pooling techniques (Section 12.2.5), other combination methods (Section 12.2.6), and empirical analyses (Section 12.2.7). Additional theoretical details can be found, for instance, in the book by (Clements and Hendry (1998), ch. 10), while Timmermann (2006) presents a detailed survey of the literature.

12.2.1 The problem of forecast combination

Let $\hat{y}_{t+h,t} = (\hat{y}_{t+h,1}, \hat{y}_{t+h,2}, \ldots, \hat{y}_{t+h,N})'$ denote an N-vector of h-step ahead forecasts for the variable y, formulated in period t. The general forecast combination problem seeks to reduce the information available in $\hat{y}_{t+h,t} \in \mathbb{R}^N$ to a lower dimensional summary measure $C(\hat{y}_{t+h,t}, \omega_c) \in \mathbb{R} \subset \mathbb{R}^N$, where ω_c are the weights associated with the combination.

We assume that the loss function only depends on the forecast error $e^c_{t+h,t} = y_{t+h} - C(\hat{y}_{t+h,t}, \omega_{t+h,t})$, i.e., $L = L\left(e^c_{t+h,t}\right)$.

The vector of optimal weights, $\omega^*_{t+h,t}$, is the solution to the following problem:

$$\omega^*_{t+h,t} = \arg \min_{\omega_{t+h,t} \in W_t} E\left[L\left(e^c_{t+h,t}(\omega_{t+h,t})\right) | \hat{y}_{t+h,t}\right] \tag{12.1}$$

Most of the times the loss function underlying the combination problem is assumed to be the mean square forecast error (MSE):

$$L\left(e^c_{t+h,t}\right) = \left(e^c_{t+h,t}\right)^2 \tag{12.2}$$

Since it is

$$E\left[\left(e^c_{t+h,t}\right)^2\right] = E^2\left[e^c_{t+h,t}\right] + var\left(e^c_{t+h,t}\right)$$

the goal of the forecast combination should be to optimally trade-off the potential increase in bias with the decrease in variance.

12.2.2 Linear forecast combination

As a starting point, we consider linear forecast combination, so that the pooled forecast is

$$\hat{y}^c_{t+h,t} = \omega_{0t+h,t} + \omega'_{t+h,t}\hat{y}_{t+h,t} \tag{12.3}$$

The optimal values of the constant and of the combination weights, $\omega^*_{0t+h,t}$ and $\omega^*_{t+h,t}$, obtained by minimizing $E\left[\left(y_{t+h} - \hat{y}^c_{t+h,t}\right)^2\right] = E\left[\left(y_{t+h} - \omega_{0t+h,t} - \omega'_{t+h,t}\hat{y}_{t+h,t}\right)^2\right]$, are:

$$\begin{cases} \omega^*_{0t+h,t} &= \mu_{yt+h,t} - \omega^{*'}_{t+h,t}\mu_{\hat{y}t+h,t} \\ \omega^*_{t+h,t} &= \Sigma^{-1}_{\hat{y}\hat{y}t+h,t}\sigma_{y\hat{y}t+h,t} \end{cases} \tag{12.4}$$

when the joint distribution of y_{t+h} and $\hat{y}_{t+h}$ is

$$\begin{pmatrix} y_{t+h} \\ \hat{y}_{t+h} \end{pmatrix} \sim \left(\begin{pmatrix} \mu_{yt+h,t} \\ \mu_{\hat{y}t+h,t} \end{pmatrix}, \begin{pmatrix} \sigma^2_{yt+h,t} & \sigma'_{y\hat{y}t+h,t} \\ \sigma_{y\hat{y}t+h,t} & \Sigma_{\hat{y}\hat{y}t+h,t} \end{pmatrix} \right) \tag{12.5}$$

Hence, more accurate forecasts that are less strongly correlated with other forecasts tend to have larger weights. Notice also that the constant corrects for any biases in the weighted forecasts. In the case where the weights are constrained to sum up to one and all the forecasts are unbiased, it is $\omega^*_{0t+h,t} = 0$. With the additional assumption that the joint distribution of y_{t+h} and $\hat{y}_{t+h}$ is Gaussian, the linear combined forecast based on the weights in (12.4) is optimal; otherwise, there might exist a non-linear combination with lower loss.

To get a better understanding of the role of forecast combination, let us consider two forecasts that give rise to errors $e_1 = y - \hat{y}_1$ and $e_2 = y - \hat{y}_2$. Assume that the individual forecasts are unbiased, so that $e_1 \sim (0, \sigma^2_1)$ and $e_2 \sim (0, \sigma^2_2)$, where $\sigma^2_1 = var(e_1)$, $\sigma^2_2 = var(e_2)$ and $\sigma_{12} = \rho_{12}\sigma_1\sigma_2$ is the covariance between e_1 and e_2 and ρ_{12} is their correlation. Suppose that the combination weights are restricted to sum to one, with weights $(\omega, 1-\omega)$. The forecast error from the combination is:

$$e^c = \omega e_1 + (1-\omega)e_2 \tag{12.6}$$

The optimal weights resulting from the formula in (12.4) are equal to:

$$\begin{cases} \omega^* & = & \frac{\sigma^2_2 - \sigma_{12}}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}} \\ 1-\omega^* & = & \frac{\sigma^2_1 - \sigma_{12}}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}} \end{cases} \tag{12.7}$$

Therefore, a greater weight is assigned to the model producing more precise forecasts. It can be easily verified that the expected squared loss associated with the optimal weights is:

$$\sigma^2_c(\omega^*) = \frac{\sigma^2_1\sigma^2_2(1-\rho^2_{12})}{\sigma^2_1 + \sigma^2_2 - 2\rho_{12}\sigma_1\sigma_2} \tag{12.8}$$

and that $\sigma^2_c(\omega^*) \leq \min(\sigma^2_1, \sigma^2_2)$. That is, we are in the presence of diversification gains, except when either σ_1 or σ_2 are equal to zero; or $\sigma_1 = \sigma_2$ and $\rho_{12} = 1$; or $\rho_{12} = \frac{\sigma_1}{\sigma_2}$.

It is instructive to compare the variance of the forecast error from the optimal combination (12.7) to the variance of the combination scheme we obtain when we choose the weights equal to the inverse of their relative mean squared errors:

$$\begin{cases} \omega^* & = & \frac{\sigma^2_2}{\sigma^2_1 + \sigma^2_2} \\ 1-\omega^* & = & \frac{\sigma^2_1}{\sigma^2_1 + \sigma^2_2} \end{cases} \tag{12.9}$$

This is a typical choice in practice, in particular when the number of forecasts under comparison, N, is large. After some algebra, we can derive the ratio of the forecast error variance under this scheme relative to its value under optimal weights:

$$\frac{\sigma^2_{inv}}{\sigma^2_c(\omega^*)} = \left(\frac{1}{1-\rho^2_{12}}\right)\left(1 - \left(\frac{2\sigma_{12}}{\sigma^2_1 + \sigma^2_2}\right)^2\right) \tag{12.10}$$

The ratio exceeds unity unless either $\sigma_1 = \sigma_2$ or $\rho_{12} = 0$.

Problems can arise when one or more of the individual forecasts are biased, the weights are constrained to sum to unity and an intercept is omitted from the combination scheme. In general, the larger the bias, the higher the probability that the combination forecast will not dominate the individual forecasts.

A further special case that deserves attention, due to its common use in empirical applications, is equally weighted combinations. We can consider the same comparison as above: we construct the ratio of the expected loss from the combination scheme with equal weights to that from the optimal combination (12.7):

$$\frac{\sigma^2_{ew}}{\sigma^2_c(\omega^*)} = \left(\frac{(\sigma^2_1 + \sigma^2_2)^2 - 4\sigma^2_{12}}{4\sigma^2_1\sigma^2_2(1-\rho^2_{12})} \right) \tag{12.11}$$

It can be shown that the ratio exceeds unity unless $\sigma_1 = \sigma_2$.

More generally, equal weights are optimal in situations with an arbitrary number of forecasts when the individual forecast errors have the same variance and identical pair-wise correlations. Moreover, combined forecasts, and in particular equally weighted combinations, seem to work well empirically also because they provide insurance against deterministic structural breaks: that is, the performance of the combined forecasts tends to be far more stable than that of the individual constituent forecasts.

One notable contribution is Hendry and Clements (2002), who examine simple mean combination forecasts when the individual models omit relevant variables and these variables are subject to out-of-sample mean shifts, which induce intercept shifts in the individual misspecified forecasting models. They show that averaging guarantees insurance and may provide dominance when the models are differentially misspecified, that when time-series are subject to deterministic shifts the average of a group of forecasts from differentially mis-specified models can outperform them, and that equal averaging may well dominate over estimated combination weights. In practice, trimmed means might be needed to exclude outlying forecasts in order to avoid that one really poor forecast worsens the combination needlessly.

12.2.3 Estimation of the combination weights

After having established the general usefulness of combining forecasts, the next step is to estimate the combination weights. It is common to estimate them in a linear model by ordinary least squares. Different versions have been proposed:

$$\begin{array}{lcl} y_{t+h} & = & \omega_{0h} + \omega'_h \hat{y}_{t+h,t} + \epsilon_{t+h} \\ y_{t+h} & = & \omega'_h \hat{y}_{t+h,t} + \epsilon_{t+h} \\ y_{t+h} & = & \omega'_h \hat{y}_{t+h,t} + \epsilon_{t+h} \text{ , such that } \omega'_h 1 = 1 \end{array} \tag{12.12}$$

The first two can be estimated by standard least squares, while the third one requires constrained least squares (see Granger and Ramanathan (1984)). When there is an intercept, we do not require the individual forecasts to be unbiased since any bias can be adjusted through the intercept term. However, not including an intercept and imposing that the weights sum to one can generate a bias, even though imposing a constraint can have the usual efficiency gain.

The relaxation of the implicit zero intercept assumption and of the summing to unity of the weights cannot worsen the within-sample mean square loss, since imposing restrictions will always increase the residual sum of squares. Diebold (1988) considers the relative merits of the restricted approach and the unrestricted regression methods from the point of view of the serial correlation properties of combined regression residuals. The predictor errors will only be serially uncorrelated in the unrestricted method if both the series to be forecasted and the forecast errors are uncorrelated. Diebold argues that the individual model errors are reasonably supposed to be white noise, so that the serial correlation in the combined prediction errors is due to the serial correlation in the series itself. This can be removed by imposing the summing-up constraint at the cost of a larger within-sample squared error. As a consequence of this, Diebold proposes estimating dynamic combining equations:

$$\begin{aligned} y_{t+h} &= \omega_{0h} + \omega_h' \hat{y}_{t+h,t} + \epsilon_{t+h} \\ \epsilon_{t+h} &= \rho \epsilon_{t+h-1} + \xi_{t+k} \end{aligned} \tag{12.13}$$

Coulson and Robbins (1993) generalize this approach by stressing the importance of testing the implicit common-factor restriction in (12.13), initially allowing more general dynamic structures. Estimation errors in the combination weights tend to be particularly large due to the difficulties in estimating the covariance matrix when the forecasts are close to collinear, and the typically short sample available. Because of this, seemingly suboptimal combination schemes such as equal-weighting have often been found to dominate other combination methods that would have been optimal in the absence of parameter estimation errors. Another possibility, as we have seen in the previous section, is to simply ignore correlations across forecast errors. This was originally advocated by Bates and Granger (1969) and Newbold and Granger (1974), who proposed combination weights that reflect the performance of each individual model relative to the performance of the average model, but ignore correlations across forecasts. Stock and Watson (1999) propose a set of combination weights that ignore correlations across forecast errors but that are based on the models' relative MSE performance raised to various powers.

The disadvantages coming from weight estimation also raise the question of whether to combine the forecasts or try to identify the best individual forecasting model. Since forecast pooling is often justified when the forecast user's information set is not the union of the information sets underlying the individual forecasts, a test of whether one forecast encompasses all information contained in another one can be useful. Chong and Hendry (1986) have developed forecast encompassing tests in the context of MSE loss function (see e.g. Marcellino (2000) for extensions of the basic procedure). Forecast encompassing concerns whether the forecasts of one model can explain the forecast errors made by another model. Generally, a test that the forecast of some model encompasses all other models can be based on a test of $\beta_2 = \cdots = \beta_N = 0$ in the regression:

$$y_{t+h} - \hat{y}_{t+h,1} = \beta_0 + \sum_{i=2}^{N} \beta_i \hat{y}_{t+h,t,i} + e_{t+h,t} \tag{12.14}$$

When the outcomes do not show a single dominant model, exploiting information by forecasts combination makes sense.

To conclude, all the considerations on estimation we have made so far are useful in a quadratic loss framework. Outside this one, the estimation of the combination weights can be based directly on the loss function (Elliot and Timmerman (2004)), and the generalized method of moments (GMM) used to estimate them.

12.2.4 Large N forecast combining methods

When the number of forecasts, N, is relatively large with respect to the number of available forecast periods, there can be considerable estimation uncertainty in the OLS weights. And in the case where N is larger than T, OLS based weight determination is no longer feasible. Therefore, in this section we discuss combination methods that can be applied in the large N context, which is becoming more and more relevant in practice due to the increased availability of alternative forecasts and forecasting methods.

Simple combination forecasts

The OLS weight estimators considered so far require, to be successful, large data samples and no outliers. When these conditions are not satisfied, simple combination schemes such as an equal weights or weights based on the inverse MSE-values offer attractive alternatives (see Sections 12.2.2 and 12.2.3). Simple combination forecasts tend to produce a value close to the center of the distribution of the panel of forecasts, and are less sensitive to outliers. Discounted MSE forecasts compute the combination forecast as a weighted

average of the individual forecasts, where the weights depend inversely on the historical performance of each individual forecast (that is, the weight on the i^{th} forecast depends inversely on its discounted MSE).

Shrinkage methods

Shrinkage forecasts entail shrinking the weights towards a value imposed a priori, which is typically equal weighting. Shrinkage methods aim to trade off bias in the combination weights against reduced parameter estimation error in estimates of the combination weights. For example, Diebold and Pauly (1990) suggest shrinkage combining weights of the form

$$\omega_{it} = \lambda\hat{\omega}_{it} + (1-\lambda)\frac{1}{n} \tag{12.15}$$

where $\hat{\omega}_{it}$ is the i^{th} estimated coefficient from a recursive OLS regression of Y^h_{t+h} on $\hat{Y}^h_{1,s+h|s}, \ldots, \hat{Y}^h_{n,s+h|s}$ for $s = T_0, \ldots, t-h$, where T_0 is the first data for the forecast combining regressions and where λ controls the amount of shrinkage towards equal weighting. Shrinkage forecasts can be interpreted as a partial implementation of Bayesian model averaging (BMA) (see Section 12.2.6).

Trimming and grouping

Rather than combining the full set of forecasts, it is often advantageous to discard the models with the worst performance (trimming). When combination weights have to be estimated, forecasts which add only marginal information should be dropped from the combination because the cost of their inclusion (that is, the increased parameter estimation error) is higher than the potential benefits of their inclusion. As an example of trimming scheme that removes poorly performing model in a step that precedes calculation of combination weights, see e.g. Granger and Jeon (2004).

Grouping or clustering of forecasts can instead be motivated by the assumption of a common factor structure underlying the forecasting models. Consider the factor model

$$\begin{array}{rcl} y_{t+h} & = & \mu_y + \beta_y' f_{t+h} + \epsilon_{yt+h} \\ \hat{y}_{t+h,t} & = & \mu_{\hat{y}} + B f_{t+h} + \epsilon_{t+h} \end{array} \tag{12.16}$$

where f_{t+h} is an $n_f \times 1$ vector of factor realizations which satisfies $E\left[f_{t+h}\epsilon_{yt+h}\right] = 0$, $E\left[f_{t+h}\epsilon'_{t+h}\right] = 0$ and $E\left[f_{t+h}f'_{t+h}\right] = \Sigma_f$. β_y is an $n_f \times 1$ vector while B is an $N \times n_f$ matrix of factor loadings. We assume that the factors have been orthogonalized. Furthermore, we also assume $E\left[\epsilon^2_{yt+h}\right] = \sigma^2_{\epsilon y}$, $E\left[\epsilon_{yt+h}\epsilon_{t+h}\right] = 0$, that all innovations are serially uncorrelated with zero mean, and the noise in the individual forecasts are idiosyncratic. Further suppose that the N forecasts can be divided into appropriate groups according to their factor loadings (clusters in the B matrix). Calculation of the optimal weights shows that pooling forecasts within each cluster and ignoring correlations across clusters does not cause a big loss, provided that the forecasts tracking the individual factors can be grouped and have similar factor exposure (see Timmermann (2006) for details on these calculations, and Section 12.2.6 for an additional discussion of factor based combination methods).

12.2.5 Time-varying parameter weights and non-linear combination methods

So far we have concentrated on schemes with linear and constant combination weights. We can generalize this approach by extending the analysis in two ways: (i) linear combinations with time-varying weights:

$$\hat{y}^c_{t+h,t} = \omega_{0t+h,t} + \omega'_{t+h,t}\hat{y}_{t+h,t} \tag{12.17}$$

or (ii) non-linear combinations with constant weights:

$$\hat{y}^{c}_{t+h,t} = C(\hat{y}_{t+h,t}, \omega) \tag{12.18}$$

where $C(.)$ is some function non-linear in the parameters, in the vector of forecasts or in both.

Time-varying parameter weights

Time-varying parameter weighting allows the weights to evolve as a stochastic process, thereby adapting to possible changes in the underlying covariances. Bates and Granger (1969) and Newbold and Granger (1974) suggested either assigning a disproportionately large weight to the model that has performed best most recently or using an adaptive updating scheme that puts more emphasis on recent performance in assigning the combination weights. Bates and Granger proposed different adapting estimation schemes based on exponential discounting or the use of rolling estimation windows. For example, a combination scheme uses rolling window of the most recent observation based on the forecasting models' relative performance (taking into account correlations across forecast errors or not). Another combination scheme uses adaptive updating, which tends to smooth the time-series evolution in the combination weights. Finally, other methods are based on exponential discounting versions of the previous schemes. An alternative to the combination schemes previously introduced is to explicitly let the combination weights evolve smoothly according to a time-varying parameter model:

$$\begin{array}{rcl} y_{t+h} & = & \breve{\omega}'_{t+h,t} z_{t+h} + \epsilon_{t+h} \\ \breve{\omega}_{t+h,t} & = & \breve{\omega}_{t+h,t-1} + \eta_{t+h} \end{array} \tag{12.19}$$

where $z_{t+h} = \begin{pmatrix} 1 & \hat{y}_{t+h,t} \end{pmatrix}'$ and $\breve{\omega}_{t+h,t} = \begin{pmatrix} \omega_{0t+h,t} & \omega'_{t+h,t} \end{pmatrix}'$. The model in (12.19) is already in state-space form, and therefore the Kalman Filter can be applied for the estimation of the weights.

Moreover, we can consider combination weights in which changes occur discretely, driven by a switching indicator I_e (see e.g. Deutsch et al. (1994)):

$$\hat{y}_{t+h} = I_{e_t \in A} \left(\omega_{01} + \omega'_1 \hat{y}_{t+h,t}\right) + \left(1 - I_{e_t \in A}\right) \left(\omega_{02} + \omega'_2 \hat{y}_{t+h,t}\right) \tag{12.20}$$

$I_{e_t \in A}$ is an indicator function which takes the value of one when $e_t \in A$ and zero otherwise, for A some pre-defined set defining the switching condition. In this case the weights can be estimated by nonlinear least squares.

Elliot and Timmerman (2004) proposed a flexible mixture model for time-variation in the combination weights. This approach supposes that the joint distribution of $\begin{pmatrix} y_{t+h} & \hat{y}'_{t+h} \end{pmatrix}$ is driven by an unobserved state variable S_{t+h} which assumes one of n_s possible values and is driven by a first-order Markov chain:

$P = Pr\left(S_{t+h} = s_{t+h} | S_t = s_t\right)$.

In this way, this approach is able to track both sudden and discrete as well as more gradual shifts in the joint distribution $\begin{pmatrix} y_{t+h} & \hat{y}'_{t+h} \end{pmatrix}$. Following the assumptions of this approach, the optimal combination weights vary as the state probabilities vary over time as a function of the arrival of new information provided that P is of rank greater than one.

Non-linear combination methods

There are two main kinds of non-linearity in forecast combinations. We can have non-linear functions of the forecasts but nevertheless have a combination that is linear in the unknown parameters, for example:

$$\hat{y}^{c}_{t+h,t} = \omega_0 + \omega' C\left(\hat{y}_{t+h,t}\right) \tag{12.21}$$

In this case non-linearity only enters in the shape of the function $C(.)$, while the parameters can be estimated by OLS.

A more general combination method considers non-linearities in the combination parameters:

$$\hat{y}^c_{t+h,t} = C\left(\hat{y}_{t+h,t}, \omega\right) \tag{12.22}$$

In this case estimation errors are expected to be even larger than in linear combination schemes; moreover, parameters are generally less robust and more sensitive to outliers than those of the linear schemes. Donaldson and Kamstra (1996) used artificial neural networks to combine volatility forecasts from a range of alternative models.

12.2.6 Other forecast combination methods

In this section we consider the combination not only of point but also of interval and probability forecasts. Next, we discuss how recent developments in Bayesian econometrics and in the large dataset econometrics can be exploited in the forecast combination context.

Combination of interval and probability distribution forecasts

So far we have focused only on combining point forecasts, however it is natural to extend the analysis to interval and probability distribution forecasts. As in the case described so far, the first question is whether it is best to use a single probability forecast or a combination of these. Suppose we are considering N distribution forecasts $f_1, \ldots, f_N$ whose joint probability with y is $P(y, f_1, \ldots, f_N)$. We can factor this probability as the product of the conditional distribution of y given $f_1, \ldots, f_N$ and the marginal distribution of the forecasts, obtaining:

$$P(y, f_1, \ldots, f_N) = P(f_1, \ldots, f_N)P(y|f_1, \ldots, f_N) \tag{12.23}$$

Clemen et al. (1995) define as extraneous a probability forecast that does not provide information about y given all the other probability density forecasts. That is, the i^{th} forecast is defined extraneous if:

$$P(y|f_1, \ldots, f_N) = P(y|f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_N) \tag{12.24}$$

In that case, probability forecast f_i does not contain useful information for forecasting y given the other $N-1$ probability forecasts. Referring back to the terminology used in previous sections, if forecast i is not extraneous, it follows that this model is not encompassed by other models. Interestingly, adding more forecasting models can lead a previously extraneous model to become non-extraneous if it contains information about the relationship between the existing $N-1$ methods and the new forecast.

Another useful concept for pairwise comparison of probability forecasts is the one of sufficiency (see Clemen et al. (1995)). If forecast 1 is sufficient for forecast 2, then forecast 1 will be of greater value to all users than forecast 2, while if either model is sufficient for the other some users will prefer model 1 and others model 2.

Combinations of probability density or distribution forecasts impose more requirements than those seen for combinations of point forecasts. Specifically, the combination must be convex with weights comprised in the zero-one interval, so that the probability forecast is never negative, and the probabilities always sums to one.

A natural combination scheme is a convex combination (also called linear opinion pool):

$$\bar{F}^c = \sum_{i=1}^{N} \omega_{t+h,t,i} F_{t+h,t,i} \tag{12.25}$$

with $0 \leq \omega_{t+h,t,i} \leq 1$ and $\sum_l imits_{i=1}^{N} \omega_{t+h,t,i} = 1$.

This scheme can be generalized by adding an extra probability forecast, $F_{t+h,t,0}$:

$$\bar{F}^c = \sum_{i=0}^{N} \omega_{t+h,t,i} F_{t+h,t,i} \tag{12.26}$$

where the weights are allowed to be negative, in particular to be in the interval $[-1, 1]$, but still restricted to sum to one.

Bayesian model averaging

Bayesian model averaging (BMA) can be thought as a Bayesian approach to combination forecasting. Also in BMA the forecast is a weighted average of the individual forecasts; however in this case the weights are computed as formal posterior probabilities that the models are correct. Moreover, the individual forecasts are model-based and are the posterior means of the variable to be forecast, conditional on the selected model. Thus BMA extends forecasts combining to a fully Bayesian setting, where the forecasts themselves are optimal Bayes forecasts, given the model. BMA were laid out by Leamer (1978), Raftery et al. (1997) and Hoeting et al. (1999).

Suppose that the conditional distribution of y_{t+h} is given by one of N models, each of them linear, so that they differ only by which subset of predictors are contained in the model. Then the predictive density for y_{t+h} is:

$$f\left(y_{t+h}|F_t\right) = \sum_{i=1}^{N} Pr\left(M_i|F_t\right) f_i\left(y_{t+h}, \Theta_i|F_t\right) \tag{12.27}$$

where Θ_i is the parameter that characterizes model i, $Pr\left(M_i|F_t\right)$ is the posterior probability of model M_i obtained from the model priors $Pr\left(M_i\right)$, $f_i\left(y_{t+h}, \Theta_i|F_t\right)$ is the predictive density of y_{t+h} and under the i^{th} model, given the information at time t.

BMA can be considered as an extension of the combination scheme proposed by Bates and Granger (1969), where the weights are determined by posterior probabilities over the models, the forecasts are posterior means, and there is no constant term (because the individual forecasts are already conditional means). Note also that the weights calculated with BMA do not account for correlations among forecasts. However, the approach is quite general and does not require the use of conjugate families of distributions.

Implementation of BMA requires choosing the prior distribution of the parameters given the model and the prior probability of the model. Given the large number of models to evaluate, it is convenient to use priors that are computationally convenient, in particular priors that impose little prior information and that lead to posteriors that are easy to evaluate quickly.

Different forecasters using the same BMA framework but different priors will produce different forecasts, and some of those forecasts will be better than others: the data can inform the choice of priors. Empirical Bayes estimation is centered on the idea of using Bayes methods with an estimated, rather than subjective, prior distribution. The empirical Bayes model consists of the regression equation for the variable to be forecasted and a specification of the priors. This approach treats the prior as an unknown distribution to be estimated. The estimation of the prior can be done either parametrically (that is, specifying a parametric prior distribution with an unknown parameter vector common to all the priors) or nonparametrically (treating the prior as an unknown distribution).

Dynamic factor models and principal component analysis

Factor analysis and principal components (PCA) are two longstanding methods for summarizing the main sources of variation and covariation among variables. Early applications of dynamic factor models (DFM) to

macroeconomic data suggest that a small number of factors can explain much of the variation of economic aggregates. If a forecaster is able to accurately estimate these factors, then he can use the estimated dynamic factors instead of the series themselves. This approach may work well because recent works suggest that large macroeconomics data sets are well described by few common dynamic factors and that the common factors can be estimated by principal components (see e.g. Stock and Watson (2002)).

The dynamic factor model is based on the assumption that the covariation among economic variables at leads and lags can be reduced to a few underlying unobserved series, named factors. The disturbances of these factors represent the aggregate shocks to the economy (for instance, demand and supply shocks).

Principal component forecast combination entails recursively computing the first few principal components (estimated common factors) of the panel of forecasts, estimating the regression of y_{t+h} onto these principal components, and forming the forecast based on this regression. This is a convenient method for allowing some estimation of factor weights and at the same time reducing the number of weights to be estimated.

Let $\hat{F}^h_{1,s}, \ldots, \hat{F}^h_{m,s}$ denote the first m estimated common factors of $\hat{y}^h_{1,s+h|s}, \ldots, \hat{y}^h_{n,s+h|s}$, for $s = T_0, \ldots, t$, computed as the first m principal components of the uncentered second moment matrix of the recursive forecasts over $s = T_0, \ldots, t$. The principal component combination forecasts are computed using the regression

$$y^h_{s+h} = \alpha_1 \hat{F}^h_{1,s} + \cdots + \alpha_m \hat{F}^h_{m,s} + v^h_{s+h} \tag{12.28}$$

where the regression coefficients $\alpha_1, \ldots, \alpha_m$ are estimated by OLS over the sample $s = T_0, \ldots, t-h$. The combined forecast is computed using the estimated weights, applied to $\hat{F}^h_{1,t}, \ldots, \hat{F}^h_{m,t}$ (see Chan, Stock and Watson (1999)).

12.2.7 Empirical analyses

A voluminous empirical literature has flourished around forecast combinations, and it is possible to draw some general conclusions from this work.

One of the most significant empirical findings on forecast combinations suggests that simple combination schemes, that is combinations that do not require the estimation of many parameters, are difficult to beat. This fact, that simple combining methods – the mean, trimmed mean and median – work well, has been called the "forecasts combining puzzle", since the theory suggests that in general it should be possible to improve upon simple combination forecasts, as we have seen.

Arithmetic averages or weights calculated as the inverse mean squared forecast error perform better than more sophisticated schemes which rely on weights that take into account the correlations of forecasts errors. The main advantage of these simple combination schemes is that their weights are known and do not need to be estimated. Therefore in many situations a simple average can achieve a substantial reduction in bias through simply averaging out individual bias and dominate forecasts based on more sophisticated techniques in terms of MSE, if due consideration is given to the effect of sampling errors and model uncertainty (see Palm and Zellner (1992)).

Consistently with the theoretical framework, the impressive performance of equal-weighted forecast combinations depends on the ratio of forecast error (not too far away from unit) and on correlation between forecast errors (not too different across pairs of models). Therefore, the performance of equal weighted combinations depends on the relative size of the variance of the forecast errors associated with different forecasting methods: equal weights perform well when we observe small differences among the variances of the different forecast errors, as pointed out by Gupta and Wilton (1987).

Moreover, as we have seen, equal weights are a good alternative when model instability renders nearly impossible a precise estimation of combination weights (c.f. Clemen and Winkler (1986)), Diebold and Pauly (1987) and Palm and Zellner (1992)).

Furthermore, the empirical application of Elliot and Timmerman (2004) confirms that in absence of asymmetry in the loss function (as in the case of MSE loss) equal-weights often outperform the optimal weights. However, under asymmetric loss, the result can be the opposite, with optimal weights conducting to smaller losses.

Many studies have found that forecasts based exclusively on the model with best in-sample performance often lead to poor out-of-sample forecasting performance, and this is particularly true for complex non-linear and time-varying models. For example, Stock and Watson (1999) consider combinations of linear and nonlinear models based on a very large set of US macroeconomic variables. What they find is a strong incentive in using forecast combination methods, particularly the average or median forecast and the forecasts with weights constructed as the inverse MSE. The dominance of combination methods holds at different forecast horizons, more specifically for one-, six-, twelve-month ahead. Furthermore, the best combination methods involve forecasts across different time-series models.

Another study conducted by Winkler and Makridakis (1983) finds that a weighted average with weights inversely proportional to the sum of squared errors or with weights that depend on the exponentially discounted sum of squared errors perform better than the best individual forecasting model or than methods which require estimation of the covariance matrix for the forecast errors.

It is also worth mentioning that including very poor models in an equal-weighted combination can substantially worsen forecasting performance (Winkler and Makridakis (1983)). Granger and Jeon (2004) recommend trimming five or ten per cent of the worst models, but even more aggressive trimming has been proposed (see Aiolfi and Favero (2005) where trimming of 80% of the models is indicated).

Shrinkage methods have performed rather well in empirical studies. Diebold and Pauly (1990) show that shrinkage systematically improves forecasting performance over individual models or methods that use least squares estimates of the combination weights. Moreover, shrinkage methods perform best when the degree of shrinkage is strong (c.f. Stock and Watson (2004)). Aiolfi and Timmermann (2004) first select models into clusters, then they pool forecasts within each cluster and estimate optimal combination weights that are shrunk towards equal weights. This combination strategy leads to better forecasting performance than simpler strategies. However, in a simulation experiment Elliot (2004) finds that although shrinkage methods always dominate least squares estimation of the combination weights, the performance of shrinkage method is sensitive to the shrinkage parameter and that it is quite difficult to accurately determine this parameter.

Note also that shrinkage and trimming work in opposite directions: while shrinkage tends to attribute similar weights to all models, trimming completely discards some of them. Therefore, if some models produce poor out-of-sample forecasts, we expect that trimming works well at a first stage while shrinkage performs poorly if the combined forecast is shrunk towards an equal-weighted average. For this reason, in many situations shrinkage is preceded by a trimming step.

Empirical evidence shows that some time-variation or adaptive adjustment in the combination weights can improve forecasting performance. Nevertheless, results in this case are mixed. In an experiment which combines forecasts from the Survey of Professional Forecasters and forecasts from simple autoregressive models, Elliot and Timmerman (2004) find evidence that suggests that the best forecast combination method allows the combination weights to vary over time in a mean-reverting manner. Stock and Watson (2004) show in their study that the combined forecasts that perform best are the time-varying parameter forecasts with very little variation, the simple mean and a trimmed mean, concluding that results for methods which allow time-variation are mixed. The time-varying parameter forecasts are not robust: sometimes they work well, sometimes quite poorly, more specifically the larger the amounts of time-variation, the less robust are the forecasts. The finding that simple combination forecasts outperform complex combination methods in empirical application is to be contextualized in a framework of widespread idiosyncratic instability in the performance of individual forecast, so that it is possible to find a stable combination of the individually unstable forecasts.

As far as Bayesian moving average is concerned, aside from the contribution by Min and Zellner (1993), which used BMA methods to combine forecasts from one linear and one nonlinear model, the other applications to economic forecasting have been quite recent. Koop and Potter (2004) focus on forecasting GDP and the

change of inflation using quarterly predictors and explore a number of different priors. They conclude that BMA produce some improvements at shorter horizons, but not at the longer ones.

Finally, the analysis in Chan, Stock and Watson (1999) does not find major gains from principal component based forecast combination.

12.3 Combined forecasting methods in practice

The main message for empirical analysis is that whether forecast combination works or not has to be evaluated on a case by case basis. Other important suggestions are the following. First, simple forecast combination methods tend to be more robust than more sophisticated alternatives, and are often at least as good or even better in terms of loss function. Among them equal weighting and weighting based on the inverse of the MSE over a training sample are particularly relevant. Second, it is important to consider forecasts that are potentially not so correlated among themselves. This suggests to consider forecasts from a variety of alternative methods, and assess combinations of different groups of forecasts. Specifically, we obtain forecasts by means of:

1. the linear method,
2. the double stochastic method,
3. the STAR method,
4. the NN method,

with several alternative model specifications within each method, differing for the choice of variable transformation, transition variable, and lag length specification. Then we compute pooled forecasts as

5. averages, with equal or MSE based weights, of linear forecasts
6. averages, with equal or MSE based weights, of nonlinear forecasts
7. averages, with equal or MSE based weights, of linear and nonlinear forecasts
8. averages, based on predictive least squares, of linear and/or nonlinear forecasts
9. averages, based on predictive least squares, of linear, nonlinear and pooled forecasts
10. medians, of linear and/or nonlinear forecasts

In order to have a comprehensive evaluation of the benefits from forecast combination, the competing forecasts are computed for a large set of PEEIs, 58 at the monthly level and 51 at the quarterly level. For each monthly indicator we consider 1-, 3-, 6-, 12-step ahead forecasts, and 1-, 2-, 4-quarter ahead forecasts for the quarterly indicators.

The resulting forecasts are assessed on the basis of four main criteria. First, the distribution over the variables of the MSE of each model. This provides an indication of the average performance of a model across variables, but also of its dispersion, which is useful as a measure of robustness. Second, the ranking of each model on the basis of the percentage of variables for which it is among the top-N models, for several values of N. This provides additional information on the robustness of the performance of each model and method. Third, the average value over all considered PEEIs of several loss functions, including the common mean absolute and mean square forecast error (MAE and MSFE, respectively). This can provide information on the overall performance of a given model or method. Finally, the identification of the best performing models according to the different loss function, indicator by indicator. This can provide information on the best models and methods for each specific PEEI. This second part is structured as follows. The dataset is illustrated in Section 12.3.1. Section 12.3.2 describes the forecasting models under investigation, and Section 12.3.3 the evaluation criteria. The results of the forecast comparison exercise are presented in Sections 12.3.4 and 12.3.5 for, respectively, monthly and quarterly time series, with some concluding remarks in Section 12.4.

12.3.1 The data

The series we analyze are extracted from the PEEI dataset (Eurostat database), and we have tried to have a fairly complete description of the economy. Actually, the monthly and quarterly series under evaluation include output variables (such as industrial production and GDP, and their components); labour market variables (such as employment, unemployment and hours worked); prices (such as consumer and producer prices); interest rates; and other miscellaneous series (including, for instance, the sentiment indicator and the business and consumer survey). The sample size typically ranges from 1995 to the end of 2007 or the first quarters of 2008.

All the series under analysis are seasonally adjusted. This can have implications for the forecasting performance of nonlinear models, see e.g. Ghysels et al. (1996) and Lee and Siklos (1997), an interesting issue for additional research in this field.

12.3.2 The forecasting methods

Following the notations of Section 12.2, the formulation of a generic forecasting model is

$$y_{t+h}^{h} = f\left(Z_t; \Theta_{ht}\right) + \epsilon_{t+h} \tag{12.29}$$

where y_t is the variable being forecasted, h indicates the forecast horizon, Z_t is a vector of predictor variables, ϵ_t is an error term, and Θ_h is a vector of parameters, possibly evolving over time. We introduce a distinction between forecasting methods and forecasting models. Forecasting methods differ for the choice of the functional form f of the relationship between y_{t+h}^{h} and Z_t. Within each method, different models are determined by the choice of the regressors Z_t and the stationarity transformation applied to y_t.

The h-step forecast is

$$\hat{y}_{t+h}^{h} = f\left(Z_t; \hat{\Theta}_{ht}\right) \tag{12.30}$$

with associated forecast error

$$e_{t+h} = y_{t+h}^{h} - \hat{y}_{t+h}^{h} \tag{12.31}$$

When y_t is treated as stationary, we use $y_{t+h}^{h} = y_{t+h}$, while if y_t is $I(1)$ then $y_{t+h}^{h} = y_{t+h} - y_t$. We present results for both cases since the choice between $I(0)$ and $I(1)$ is less clear cut with time-varying models. Moreover, we also consider a pre-test forecast where the decision on the stationarity of y_t is based on a unit root test, which often improves the forecasting performance in the case of linear models, see e.g. Diebold and Kilian (2000). In particular, we use the DF-GLS statistics of Elliot et al. (1996), which performed best in the simulation experiments in Stock (1996). Note that $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$, independently of whether y_t is treated as stationary or not, so that forecast errors from the three different cases (stationary, $I(1)$ and pre-test) are directly comparable.

We focus on 1-, 3-, 6- and 12-step ahead forecasts for the monthly variables, and 1-, 2-, and 4-step ahead forecasts for the quarterly variables. Partly due to the rather short sample available, longer forecast horizons result in very uncertain forecasts. When the forecast horizon h is larger than one, the "h-step ahead projection" approach in (12.29), also called dynamic estimation (e.g. Clements and Hendry (1996)), differs from the standard approach of estimating a one-step ahead model, and iterate it forward to obtain h-step ahead predictions. The h-step ahead projection approach has two main advantages in this context. First, the impact of specification errors in the one-step ahead model can be reduced by using the same horizon for estimation as for forecasting. Second, simulation methods are not required to obtain forecasts from non-linear models. The resulting forecasts could be slightly less efficient, see e.g. Granger and Terasvirta (1993) (Ch.8) and

Marcellino et al. (2006), but the computational savings in our pseudo real time exercise with many series and models are substantial.

It is also worth noting that a few forecast errors from the non-linear and time-varying methods can be very large. This is due to problems in the estimation of these models, as discussed in Section 12.2, because of multiple local optima and the rather short estimation sample available. In order not to bias the comparison against these methods and in favour of pooling, we automatically trim the forecasts. In particular, when the absolute value of a forecasted change is larger than any previously observed change, a no change forecast is used. This is close to what a real forecaster would do when working with these models in real time.

Let us now list the methods and models we compare, and briefly discuss their main characteristics and estimation issues. More details can be found in Marcellino (2004), Stock and Watson (1999), and Timmermann (2006).

Linear methods

Autoregression (AR). Though very simple, these models have performed rather well in forecast comparison exercises, see e.g. Meese and Geweke (1984), or Marcellino et al. (2003) for the Euro area. The f function in (12.29) is linear, and Z_t includes lags of the y_t variable and a deterministic component. The latter can be either a constant or also a linear trend. The lag length is either fixed at 4, or it is chosen by AIC or BIC with a maximum of 6 lags. Given that the y_t variable can be treated as stationary, $I(1)$, or pre-tested for unit roots, overall we have 18 models in this class.

Exponential smoothing (ES). Makridakis et al. (1982) found this method to perform rather well in practice even though, from a theoretical point of view, it is optimal in the mean square forecast error sense only when the underlying process follows a particular ARMA structure, see e.g. Granger and Newbold (1986) (Ch.5). We consider both single and double exponential smoothing, which are usually adopted for, respectively, stationary and trending series. Estimation of the parameters is conducted by means of (recursive) non-linear least squares (see e.g. Tiao and Xu (1993)). The third model in this class is given by a combination of the single and double models, based on the outcome of the unit root test.

Non-linear methods

Time-varying autoregression (TVAR). Following Nyblom (1989), we let the parameters of the AR model evolve according to the following multivariate random walk model:

$$\Theta_{ht} = \Theta_{ht-1} + u_{ht}\ ,\ u_{ht} \sim iid\left(O, \lambda^2\sigma^2 Q\right) \tag{12.32}$$

where σ^2 is the variance of the error term ϵ in (12.29), $Q = \left(E\left[Z_t Z_t'\right]\right)^{-1}$, and we inspect several values of λ: 0 (no variation), 0.0025, 0.005, 0.0075, 0.01, 0.015, or 0.020. We consider first a specification with a constant, 3 lags and $\lambda = 0.005$, and then we allow for AIC or BIC selection of the number of lags (1, 3 or 6) jointly with the value of λ. In each case, y_t can be either stationary, or $I(1)$ or pre-tested, so that we have a total of 9 TVAR models. The models are estimated by the Kalman filter.

Logistic smooth transition autoregression (LSTAR). The generic LSTAR model can be written as

$$y_{t+h}^h = \alpha'\varsigma_t + d_t\beta'\varsigma_t + \epsilon_{t+h}, \tag{12.33}$$

where

$$d_t = \frac{1}{1 + e^{\gamma_0 + \gamma_1 \varsigma_t}}$$

and $\varsigma_t = (1, y_t, y_{t-1}, \ldots, y_{t-p+1})$ if y_t is treated as stationary or $\varsigma_t = (1, \Delta y_t, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1})$ if y_t is $I(1)$. The smoothing parameters γ_1 regulate the shape of parameter change over time. When $\gamma_1 = 0$ the

model becomes linear, while for large values of γ_1 the model tends to a self-exciting threshold model, see e.g. Granger and Terasvirta (1993), Terasvirta (1998) for details. For models specified in levels we consider the following choices for the threshold variable in d_t: $\varsigma_t = y_t$, $\varsigma_t = y_{t-2}$, $\varsigma_t = y_{t-5}$, $\varsigma_t = y_t - y_{t-6}$, $\varsigma_t = y_t - y_{t-12}$. For differenced variables, it can be $\varsigma_t = \Delta y_t$, $\varsigma_t = \Delta y_{t-2}$, $\varsigma_t = \Delta y_{t-5}$, $\varsigma_t = y_t - y_{t-6}$, $\varsigma_t = y_t - y_{t-12}$. In each case the lag length of the model was either 1 or 3 or 6. We report results for the following models: 3 lags and $\varsigma_t = y_t$ (or $\varsigma_t = \Delta y_t$ for the $I(1)$ case); 3 lags and $\varsigma_t = y_t - y_{t-6}$; AIC or BIC selection of both the number of lags and the specification of ς_t. In each case, y_t can be either stationary, or $I(1)$ or pre-tested, so that overall there are 12 LSTAR models. Estimation is carried out by (recursive) non-linear least squares, using an optimizer developed by Stock and Watson (1999).

Artificial neural network (ANN). ANN models can provide a valid approximation for the generating mechanism of a vast class of non-linear processes, see e.g. Hornik et al. (1989), and Swanson and White (1997) for their use as forecasting devices. The single layer feedforward neural network model with n_1 hidden units (and a linear component) is specified as:

$$y_{t+h}^h = \beta_0' \varsigma_t + \sum_{i=1}^{n_1} \gamma_{1i} g\left(\beta_{1i}' \varsigma_t\right) + \epsilon_{t+h} \tag{12.34}$$

where $g(x)$ is the logistic function, $g(x) = \frac{1}{1+e^x}$. Even more flexibility can be obtained with the double layer feedforward neural network with n_1 and n_2 hidden units:

$$y_{t+h}^h = \beta_0' \varsigma_t + \sum_{j=1}^{n_2} \gamma_{2j} g\left(\sum_{i=1}^{n_1} \beta_{2ji} g\left(\beta_{1i}' \varsigma_t\right)\right) + \epsilon_{t+h} \tag{12.35}$$

We report results for the following specifications: $n_1 = 2$, $n_2 = 0$, $p = 3$ (recall that p is number of lags in ς_t); $n_1 = 2$, $n_2 = 1$, $p = 3$; $n_1 = 2$, $n_2 = 2$, $p = 3$; AIC or BIC selection with $n_1 = (1, 2, 3)$, $n_2 = (1, 2$ with $n_1 = 2)$, $p = (1, 3)$. For each case y_t can be either stationary, or $I(1)$ or pre-tested, which yields a total of 15 ANN models. The models are estimated by (recursive) non-linear least squares, using an algorithm developed by Stock and Watson (1999).

No-change

No change (NC). The random walk based forecast is simply $\hat{y}_{t+h} = y_t$. Not withstanding its simplicity, in a few cases it was found to outperform even forecasts from large-scale structural models, see e.g. Artis and Marcellino (2001).

Pooling procedures

Linear combination forecasts (C). These forecasts are weighted averages of those described so far:

$$\hat{y}_{t+h} = \sum_{m=1}^{M} k_{m,h,t} \hat{y}_{t+h,m} \text{ , } k_{m,h,t} = \left(\frac{1}{msfe_{m,h,t}}\right)^w / \sum_{j=1}^{M} \left(\frac{1}{msfe_{j,h,t}}\right)^w \tag{12.36}$$

where m indexes the models, $k_{m,h,t}$ denotes the weights, and $msfe$ indicates the mean square forecast error. As discussed in Section 12.2, Bates and Granger (1969) showed that the weighting scheme that minimizes the $msfe$ of the pooled forecasts involves the covariance matrix of all the forecast errors, which is unfeasible in our case because M is very large. Hence, following their suggestion, the weight of a model is simply chosen as inversely proportional to its $msfe$, which is equivalent to setting $w = 1$ in equation (12.36). We also consider the cases $w = 0$, equal weight for each forecast, and $w = 5$, more weight for the best performing models. Moreover, we analyze separately pooling the linear models only, the non-linear models only, and all the models. Thus, overall we have 9 linear combination forecasts.

Table 12.1: Forecasting models under evaluation

!htp

A. Linear methods	
ARF(X,Y,Z)	Autoregressive models (18 models) X = C (const.) or T (trend) Y = 0 (stationary), 1 (I(1)), P (pre-test) Z = 4 (4 lags), a (AIC), b (BIC)
EX(X)	Exponential smoothing (3 models) X = 1 (single), 2 (double), P (pre-test)
B. Non-linear methods	
ARTVF(X,Y,Z)	Time-varying AR models (9 models) X = C (const.) Y = 0 (stationary), 1 (I(1)), P (pre-test) Z = 3 (3 lags), a (AIC), b (BIC)
LS(X,Y,Z)	Logistic smooth transition (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) Y = transition variable, 10 ($\varsigma_t = y_t$), 06 ($\varsigma_t = y_t - y_{t-6}$) Z = 3 (p, lag length)
LSF(X,W)	Logistic smooth transition (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) W = a (AIC on transition variable and p), b (BIC)
AN(X,Y,Z,W)	Artificial neural network models (9 models) X = 0 (stationary), 1 (I(1)), P (pre-test) Y = 2 (n_1) Z = 0, 1, 2 (n_2) W = 3 (p, lag length)
ANF(X,S)	Artificial neural network models (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) S = a (AIC on n_1, n_2, p), b (BIC)
C. No Change	
NOCHANGE	No change forecast (1 model)
D. Pooling	
C(X,Y)999	Linear combination (9 forecasts) X = 1 (combine A,B,C), 2 (A only), 3 (B only) Y = 0, 1, 5 (weight, w in equation (12.36))
M(X)999	Median combination (3 forecasts) X = 1 (combine A,B,C), 2 (A only), 3 (B only)
P(X)999	Predictive least square combination (4 forecasts) X = 1 (combine A,B,C), 2 (A only), 3 (B only), A (A,B,C,D)

Median combination forecasts (M). These are the median forecasts from a set of models, and are computed because with non-Gaussian forecast errors linear combinations of the forecasts are no longer necessarily optimal. As in the previous method, we distinguish among three groups of models: linear, non-linear, and all models. Thus, we have 3 median combination forecasts.

Predictive least squares combination forecasts (PLS). In this approach, described in more details in Section 12.2, the model is selected on the basis of its past forecasting performance over a certain period, which for us is one year. Thus, the model that produced the lowest $msfe$ over the past year is used as the forecasting model, and the choice is recursively updated each month over the forecast period. We compute 4 of these forecasts, which differ for the set of models compared: all models, all linear models, all non-linear models, all models plus the linear and the median combination forecasts.

The 58 models and the 16 pooling procedures to be used in the forecast comparison exercise are summarized in Table 12.1.

12.3.3 Forecast Evaluation

The evaluation of the relative forecasting performance of the $M = 74$ models for the $N = 58$ monthly time series (or $N = 51$ quarterly variables) in the dataset requires a careful choice of the loss function.

As a starting point, for variable n and forecasting method m, let us define the loss function as

$$Loss^{h}_{n,m} = \frac{1}{T-h}\sum_{t=1}^{T-h} |e_{t+h,n,m}|^{\rho} \tag{12.37}$$

where e_{t+h} is the h-step ahead forecast error, and ρ can be equal to 1, 1.5, 2, 2.5 or 3. The values $\rho = 1$ and $\rho = 2$ correspond to the familiar choices of, respectively, the mean absolute and the mean square forecast error (MSE and MAE) as the loss function. The higher the value of ρ, the larger the relative importance of large forecast errors. It is important to consider different values of ρ since some nonlinear models can produce extreme forecast errors just in a few periods, which could be assigned a different weight by different evaluators.

The loss function in (12.37) will form the basis for a set of alternative criteria to assess the relative performance of the 74 models and pooling procedures under analysis. As a first evaluation criterion, we set $\rho = 2$ in (12.37), namely, we compute the MSE for each of the 74 models and 58 variables, and then rescale the MSEs by the MSE of a benchmark AR(4) model with a constant, specified in levels. In formulae, the relative MSE (RelMSE) of model j for variable m is:

$$rmsfe^{h}_{j-AR4,m} = \sum_{t=1}^{T-h} e^{2}_{j,t+h,m} / \sum_{t=1}^{T-h} e^{2}_{AR4,t+h,m} \tag{12.38}$$

Such a rescaling makes the RelMSEs comparable across variables. Then, for each model, we calculate the empirical distribution of the RelMSE over all the variables under analysis, and we report the mean of the distribution and some percentiles.

The empirical distribution of the RelMSE criterion provides a substantial amount of information. In particular, its mean is directly related to the average performance of each model for all the variables under analysis, and therefore it is useful for an assessment of the overall performance of each forecasting model and method. In addition, the bottom and top percentiles are useful to judge whether the distribution for a given model is fairly concentrated around the mean value, or rather disperse. In the former case, it can be expected that the average forecasting performance of a model will be similar to that for any of the variables under analysis. In the latter case, there can be variables for which the model performs very well, but also other for which it performs very poorly, and therefore a more disaggregate analysis is required.

As a second evaluation criterion, we compute the fraction of variables for which a given model is in the top-N models in terms of MSE, for $N = 1, 5, 10, 15$ and 20. For example, we want to know for what fraction of series a simple pooled forecast is in the set of the 15 models with lowest MSE. Based on this criterion, the preferred forecasting models should be in the top-N models for as large a fraction of series as possible. This criterion provides complementary information with respect to the distribution of RelMSE, since the latter does not compare across models but only each model with respect to the benchmark.

As a third criterion, we compute the following average loss function for each model m:

$$Loss^{h}_{m} = \frac{1}{N}\sum_{n=1}^{N} \frac{Loss^{h}_{n,m}}{Loss^{h}_{n,1}} \tag{12.39}$$

for different values of ρ (1, 1.5, 2, 2.5 or 3). For a given model, equation (12.39) is a weighted average of the loss for each variable, with weights given by the inverse of the loss of the benchmark AR(4) forecast. We then

rank the models based on (12.39), with models with lower loss being preferred. Notice that the ranking based on average MSE ($\rho = 2$in (12.39)) could be also obtained by comparing the mean of the empirical distribution of the relative MSE. However, with (12.39) we can also check whether this ranking is robust to changes in the evaluation function, in particular whether assigning more or less weight to large forecast errors changes the ranking.

All the three criteria we have mentioned treat all the variables in the same way. However, as we have mentioned in Section 12.2, the literature on forecast pooling has emphasized that pooling could be particularly helpful in the presence of parameter instability. To assess whether this is the case for the PEEIs, we have implemented a set of parameter instability tests, and grouped those PEEIs for which at least one test rejects the null hypothesis of stable parameters at the 1% level. The tests we have used are listed in the note to Table 12.2. For example, we will see that stability is rejected for 21 out of the 58 monthly PEEIs. We have then repeated the evaluation focusing only on these 21 PEEIs.

Finally, we have considered a very disaggregate analysis. For each of the $N = 58$ variables, 4 and 3 forecast horizons (h) for, respectively, monthly and quarterly PEEIs, and $M = 74$ models, we have computed

$$Loss^{h}_{n,m} = \frac{Loss^{h}_{n,m}}{Loss^{h}_{n,1}} \tag{12.40}$$

for different values of ρ (1, 1.5, 2, 2.5 or 3), which is the equivalent of (12.39) but without averaging across variables. Hence, for each of the 58 monthly and 51 quarterly PEEIs under analysis, we have a ranking of all the 74 models and pooling procedures under evaluation, for each forecast horizon and choice of loss function. In the next two sections we will focus on the results emerging from the first three evaluation criteria (empirical distribution of RelMSE, top-N models, and average loss) for, respectively, monthly and quarterly PEEIs.

12.3.4 Empirical results for monthly PEEIs

In the first subsection we provide a detailed evaluation of the forecasting performance of the alternative models for the 58 monthly PEEIs, while in the second subsection we restrict the attention to those PEEIs whose generating mechanism is likely unstable over the period under analysis. The forecasting performance is in general evaluated over the period 2001:1-2008:6, which includes 90 months. The tables of results are available upon request to the editor.

An evaluation of all the monthly PEEIs

Let us start our evaluation of the relative forecasting performance of the 74 linear, nonlinear and pooling methods looking at the empirical distribution of the relative MSE. Nine main findings emerge. First, based on the mean of the distribution, it is difficult for any linear model to beat the benchmark AR(4) specification in levels with a constant and no trend. The only sizeable gains, about 13% in terms of lower MSE, can be obtained for $h = 6$ by imposing a unit root and selecting the lag length with the AIC specification (ARFT1a). Second, among the nonlinear and time-varying models, the Neural Networks (AN) are the worst, while time-varying AR models (ARTV) and smooth transition AR models (LS) are broadly comparable, again based on the mean of the empirical distribution. However, the benchmark is rarely beaten and never by a substantial amount. Third, both exponential smoothing and a pure random walk forecast never beat the benchmark, notwithstanding the short sample size. Fourth, and perhaps most interesting from our perspective, there seem to be some gains from forecast pooling. In particular, putting a larger weight on the best performing models over the training period ($w = 5$ in equation (12.36)), and combining all linear, nonlinear and no change models yields gains of about 5% for $h = 1$, 7% for $h = 3$, 20% for $h = 6$, but a loss of about 8% for $h = 12$. Fifth, forecast combination is generally preferable for the monthly PEEIs to either choosing the median forecast or selecting the best forecast based on the performance over the past 12 months. Sixth, if we now rank the models based on the median of the relative MSE across variables rather than the mean, it turns out that for

the linear models that are minor differences, while for the nonlinear and time-varying models the median is systematically smaller than the mean, even though in general still larger than the median for the benchmark AR(4) model. This finding suggests that the nonlinear and time-varying models perform very poorly for some variables, and this deteriorates their average performance. Seventh, looking again at the median but this time focusing on the combined forecasts, we find in general the opposite result, namely, the median across variables is higher than the mean. This indicates that the fraction of variables for which forecast combination works badly is fairly small. Eight, to provide additional evidence on the previous two comments, we can look at the 90% percentiles of the empirical distribution (with qualitatively similar comments for the 98% percentiles). Indeed, it emerges that for the combined forecasts the 90% percentile is about only 10% higher than the benchmark when $h < 12$, while for the nonlinear and time-varying forecasts often the values are 40-120% worse than the benchmark. Ninth, on the other hand, the 10% percentile is in most cases smaller than one for the nonlinear and time-varying models, which implies that they can beat the linear AR(4) benchmark at least for a fraction of the PEEIs under evaluation. Actually, for some PEEIs, the linear benchmark is indeed outperformed, at least for $h = 1$. In summary, on average over all the 58 monthly PEEIs it is difficult to beat a linear AR(4) benchmark, but pooling all the models can yield some gains for $h = 1, 3, 6$. For specific PEEIs the nonlinear or time-varying models can be either very good or very bad, while the performance of pooling appears to be more stable across variables, and this robustness feature can be relevant in an institutional contest.

An additional evidence on the robustness of forecast combination is the fraction of variables for which a given model is in the set of the N models with lowest MSE, for different values of N. If we look, for example, at $h = 1$ and $N = 1$, i.e., the best models for one month ahead forecasts of each indicator, it turns out that forecast combination is best for only 8% of the 58 monthly PEEIs, versus 27% of linear AR models, 21% of AR models with time-varying parameters, and even 17% of neural network models. However, if we look at the fraction of PEEIs for which forecast combination is among the best 20 models for $h = 1$, we get a value of 91% in the case of the combination of linear, nonlinear and no change models, which is the highest among the 74 forecasting methods undes consideration. Therefore, forecast combination is not necessarily the best forecasting tool for each indicator, but it is systematically good. On the contrary, complex specific models, such as specific AR with time-varying parameters or neural networks, can perform very well for a few indicators, but very badly for most of the other indicators.

As a further check of the robustness of the combination forecasts, we have computed the average loss function in (12.39) for different values of ρ (1, 1.5, 2, 2.5 or 3), where we remind that $\rho = 1$ corresponds to MAE and $\rho = 2$ to MSE. It turns out that combination forecasts are always the best performers for $h < 12$, with linear AR models preferred for $h = 12$. Among the nonlinear and time-varying specifications, the best ones are the time-varying AR models, but they are typically not in the top 20 models. Therefore, the good performance of the combination forecasts on average for the PEEIs is also robust to the choice of the loss function.

Table 12.2: Time Varying Parameter Tests Monthly Series. NY is Nyblom (1989) test. KP1 and KP2 are the Ploeberger and Kramer (1992) supremum and mean square tests. BP is Breusch and Pagan (1979) Lagrange multiplier test. QLR is Quandt (1960) supremum F-test. MW is Andrews and Ploeberger (1994) mean F-test. EW is Andrews and Ploeberger (1994) average exponential F-test. Test_r is robust version based on HAC variance covariance matrix. *, **, and *** indicate significance at, respectively, 10%, 5% and 1% level.

Series	Sample		Tests Lags	NY	NYr	KP1	KP2	BP	QLR	MW	EW	Break	QLRr	MWr	Ewr
CBRO	1992:11	2007:12	5	0,99	0,97	0,72	0,07	2,94	12,48	7,99	4,67	81,00	11,63	8,53	4,60
GBY	1990:01	2007:12	4	0,49	0,46	0,81	0,08	4,40	17,31*	5,27	4,74	57,00	12,26	5,51	3,41
I2000EPI	1993:01	2007:12	6	0,71	0,93	0,79	0,15	9,85	35,10***	8,72	12,79***	79,00	18,33	10,15*	6,19
I2000HWI	1995:01	2007:12	2	2,47***	2,69***	1,35**	0,69**	5,14*	26,48***	20,23***	10,97***	117,00	14,86**	11,18***	5,98***
I2000IP	1990:01	2007:12	5	1,00	1,41	0,81	0,21	20,33***	12,56	8,01	4,77	79,00	13,49	8,65	4,89
I2000WSI	1995:01	2007:12	5	1,20	0,91	1,07	0,35*	7,14	29,49***	15,01***	12,77***	44,00	13,42	9,25*	5,16
IPGR1	1990:01	2007:12	6	1,62	1,60	0,79	0,09	9,63	22,80**	10,64*	7,74**	32,00	17,75	9,43	5,89
IPGR12	1991:01	2007:12	6	1,35	1,51	0,86	0,15	3,36	17,02	9,05	6,28	29,00	15,22	8,83	5,67
MF3M	1990:01	2007:12	4	0,62	0,44	1,05	0,10	34,77***	29,77***	4,52	10,46***	31,00	9,79	4,67	2,86
NEXGBP	1994:01	2007:12	1	0,52	0,57	0,94	0,11	0,64	32,84***	6,34***	12,16***	26,00	13,75**	6,02**	4,06**
NEXJPY	1994:01	2007:12	5	1,04	0,73	0,76	0,12	6,36	12,27	7,60	4,63	96,00	11,50	6,82	4,24
NEXUSD	1994:01	2007:12	3	0,91	0,94	1,65***	0,78***	0,40	15,67*	7,85**	5,54**	92,00	16,47**	8,11**	5,68**
PPIGR1	1990:01	2007:12	3	0,67	0,83	0,97	0,21	12,80***	10,48	5,32	3,25	164,00	12,06	5,98	4,08
PPIGR12	1990:01	2007:12	6	1,15	1,12	1,20*	0,30	8,11	25,74**	8,65	8,50**	164,00	16,37	7,13	5,20
RETRADE1	1995:02	2007:12	3	0,61	0,64	1,17	0,27	6,11	8,20	3,89	2,39	74,00	7,73	4,78	2,77
RETRADE12	1996:01	2007:12	4	0,91	0,78	1,18	0,24	6,12	20,94**	7,39	7,03**	62,00	15,78	7,25	5,27*
RETRADED12	1996:01	2007:12	4	1,20	0,96	0,95	0,34*	1,47	15,10	8,72**	5,55*	118,00	12,77	7,38	4,47
RETRATED1	1995:02	2007:12	3	0,61	0,61	0,87	0,27	3,28	6,19	3,37	1,95	126,00	6,77	3,65	2,09
RT1_SA_C_D_E_IS_IP_EA13	1990:02	2007:12	6	1,63	1,61	0,78	0,08	10,04	23,15**	10,58*	7,70*	31,00	17,77	9,37	5,81
RT1_SA_C_D_E_IS_PPI_EA13	1990:01	2007:12	3	0,67	0,83	0,97	0,21	12,80***	10,48	5,32	3,25	164,00	12,06	5,98	4,08
RT1_SA_CAPITAL_IS_IP_EA1	1990:02	2007:12	6	1,06	0,99	0,99	0,17	2,61	22,24**	6,84	6,75	31,00	16,83	6,65	5,21
RT1_SA_CAPITAL_IS_PPI_EA	1990:02	2007:12	6	1,06	1,22	0,89	0,14	1,04	12,50	6,66	4,00	31,00	12,96	7,76	4,51
RT1_SA_CONS_DUR_IS_IP_EA	1990:02	2007:12	3	0,33	0,29	0,80	0,10	3,37	5,72	1,96	1,17	125,00	5,47	1,81	1,07
RT1_SA_CONS_DUR_IS_PPI_E	1991:02	2007:12	4	0,40	0,41	0,92	0,12	3,26	11,57	3,34	2,38	151,00	8,61	3,17	2,00
RT1_SA_CONS_NONDUR_IS_IP	1990:02	2007:12	2	0,37	0,45	0,78	0,14	0,96	5,17	2,51	1,45	155,00	6,19	2,92	1,76
RT1_SA_CONS_NONDUR_IS_PP	1990:01	2007:12	3	0,63	0,58	0,65	0,10	15,02***	8,39	3,99	2,34	84,00	10,72	5,00	3,16
RT1_SA_D_IS_IP_EA13	1990:02	2007:12	6	1,49	1,51	0,74	0,08	1,97	18,95	9,78	6,00	31,00	15,47	9,06	5,20
RT1_SA_D_IS_PPI_EA13	1990:01	2007:12	1	0,57	0,62*	0,99	0,20	0,07	17,10***	4,50**	5,10***	160,00	9,41	3,37	2,46
RT1_SA_ENERGY_IS_IP_EA13	1990:02	2007:12	5	1,23	1,06	0,48	0,05	28,74***	17,23	7,40	5,77	64,00	12,26	6,16	3,78
RT1_SA_ENERGY_IS_PPI_EA1	1990:02	2007:12	1	0,36	0,40	1,09	0,25	0,17	7,75	2,70	1,81	163,00	6,50	2,35	1,50
RT1_SA_INTER_IS_IP_EA13	1990:02	2007:12	3	0,73	0,57	0,86	0,10	2,39	9,54	5,11	2,87	53,00	8,50	4,81	2,84
RT1_SA_INTER_IS_PPI_EA13	1990:02	2007:12	4	0,46	0,50	0,86	0,18	20,58***	11,22	4,43	3,38	38,00	8,72	3,98	2,50
SENTIND	1986:12	2007:12	4	0,73	0,51	0,56	0,06	40,07***	22,69***	5,41	6,39**	174,00	12,39	4,07	3,00
URATE	1993:01	2007:12	6	0,63	1,20	0,89	0,23	37,12***	20,94*	7,94	7,23*	56,00	29,53***	12,22**	11,48***
RT1_NSA_CP_HI00_EA13	1996:02	2007:12	1	0,55	0,57	1,16	0,41*	2,97*	7,40	3,25	2,05	35,00	10,23*	3,84*	2,90*
RT1_NSA_CP_HI00XE_EA13	1996:02	2007:12	6	0,89	1,00	1,26*	0,25	19,11***	13,49	6,67	4,07	66,00	13,57	8,52	4,74
RT1_NSA_CP_HI00XEF_EA13	1996:02	2007:12	6	0,61	0,64	1,24*	0,21	5,30	12,14	5,13	3,61	46,00	12,07	6,35	4,11
RT1_NSA_CP_HI00XEFU_EA13	1996:02	2007:12	6	1,06	1,16	1,14	0,23	9,68	19,60*	7,61	6,89*	108,00	16,00	8,99	5,78
RT1_NSA_CP_HI00XES_EA13	1996:02	2007:12	6	0,98	1,16	1,10	0,21	12,78**	20,04*	7,29	6,92*	108,00	16,33	8,79	5,91
RT1_NSA_CP_HI00XTB_EA13	1996:02	2007:12	1	0,59	0,58	1,18	0,44*	0,64	7,04	3,54	2,08	46,00	7,98	3,72*	2,30
RT1_NSA_CP_HI01_EA13	1996:02	2007:12	2	0,41	0,40	0,86	0,19	12,24***	8,00	2,76	1,94	42,00	6,24	2,62	1,61
RT1_NSA_CP_HI02_EA13	1996:02	2007:12	1	0,33	0,34	1,02	0,29	0,01	4,44	2,05	1,22	54,00	5,52	2,73	1,63
RT1_NSA_CP_HI03_EA13	1996:02	2007:12	4	0,80	0,79	1,53**	0,35*	5,41	18,63**	5,64	6,14**	53,00	14,79	6,57	4,90
RT1_NSA_CP_HI04_EA13	1996:02	2007:12	1	0,40	0,41	1,13	0,37*	0,40	6,44	2,72	1,72	93,00	6,46	2,71	1,74
RT1_NSA_CP_HI05_EA13	1996:02	2007:12	5	1,22	1,07	0,67	0,11	1,75	16,03	7,84	5,59	92,00	12,89	7,88	4,67
RT1_NSA_CP_HI06_EA13	1996:02	2007:12	1	0,14	0,23	0,84	0,14	0,02	4,50	1,07	0,72	101,00	7,01	2,05	1,66
RT1_NSA_CP_HI07_EA13	1996:02	2007:12	2	0,61	0,69	0,82	0,19	1,76	9,15	4,03	2,45	57,00	10,01	5,22*	3,13
RT1_NSA_CP_HI08_EA13	1996:02	2007:12	3	0,65	0,73	0,70	0,11	1,47	11,06	5,36	3,27	43,00	7,32	5,18	2,76
RT1_NSA_CP_HI09_EA13	1996:02	2007:12	6	0,53	0,76	1,05	0,21	3,52	14,14	5,16	3,74	68,00	12,33	6,95	4,18
RT1_NSA_CP_HI10_EA13	1996:02	2007:12	1	0,68*	0,82**	1,46**	0,66**	0,00	20,95***	4,30*	5,95***	117,00	7,26	4,63**	2,55
RT1_NSA_CP_HI11_EA13	1996:02	2007:12	6	1,09	0,94	0,78	0,13	4,44	31,58***	8,95	11,80***	57,00	17,77	7,77	5,97
RT1_NSA_CP_HI12_EA13	1996:02	2007:12	5	0,69	0,68	0,82	0,13	3,60	9,99	6,38	3,52	31,00	11,93	7,52	4,29
RT1_NSA_CP_HIE_EA13	1996:02	2007:12	1	0,15	0,16	0,69	0,11	0,03	3,04	0,99	0,55	31,00	6,74	1,43	1,11
RT1_NSA_CP_HIF_EA13	1996:02	2007:12	2	0,63	0,60	1,09	0,26	12,71***	9,92	4,09	2,71	42,00	9,17	3,78	2,42
RT1_NSA_CP_HIFU_EA13	1996:02	2007:12	1	0,37	0,27	0,71	0,09	9,12***	8,89	2,59	1,80	66,00	6,04	2,44	1,51
RT1_NSA_CP_HIG_EA13	1996:02	2007:12	4	0,88	1,13	1,19*	0,41*	3,53	9,88	5,33	3,16	34,00	18,33**	7,63	6,40**
RT1_NSA_CP_HIIG_EA13	1996:02	2007:12	2	0,27	0,37	0,85	0,22	2,09	4,58	1,74	1,05	31,00	8,72	2,48	2,00
RT1_NSA_CP_HIIGXE_EA13	1996:02	2007:12	1	0,16	0,12	0,64	0,07	2,87*	3,50	1,01	0,60	66,00	2,88	1,09	0,61

An evaluation of the unstable monthly PEEIs

Table 12.2 reports a battery of tests for parameter constancy applied to each of the 58 monthly PEEIs, highlighting those for which at least one of the tests rejects at the 1% level. We have chosen such a low level for the test to partly control for the fact that we are applying a set of tests to a large number of indicators, and therefore some of them would be ranked as unstable as a consequence of type I errors. It turns out that 21 out of the 58 PEEIs are classified as unstable, and we now repeat the forecast assessment exercise for this subset of indicators. We would expect a better performance of nonlinear and time-varying models in this case, and also of forecast pooling techniques.

Indeed, the overall performance of the nonlinear and time-varying models improves, but in general not enough to beat the benchmark AR model. The performance of the combination forecasts also improves, the gains with respect to the full set of PEEIs case are small for $h < 12$, but they are larger for $h = 12$ and now the benchmark is beaten in most cases also for h=12. These results are based on the mean of the empirical distribution of the relative MSE, but using the median does not change the overall picture. In addition, the values of the 90% percentile of the distribution suggest that the fraction of series for which the nonlinear and time-varying models are very bad decreases, in particular the time-varying AR and Smooth Transition models.

Finally, focusing on the performance of the forecast combination, which is our main method of interest, it turns out that the combination of linear, nonlinear and no change models is rarely the best forecast for each indicator, but is in the top-20 best forecasts for a large fraction of the indicators. As a consequence, when its performance is evaluated on average across all the unstable PEEIs, it is systematically ranked either first or in the top places, independently by the choice of the loss function. It is also worth mentioning that the relative ranking of the nonlinear and time-varying models in general improves with respect to the previous section, particularly so for the time-varying AR models.

In summary, the empirical results are in line with the theoretical expectations: the performance of the more complex forecasting models improves, but in general not enough to beat the benchmark, as well as that of the forecast combination, which is still better than the benchmark for $h = 1, 3, 6$ and now also for $h = 12$.

12.3.5 Empirical results for quarterly PEEIs

The forecasting performance of the 74 linear, nonlinear and pooling methods when applied to the 51 quarterly PEEIs is evaluated in general over the period 2000:4-2008:2, which includes 31 quarters.

Five main findings emerge from the empirical distribution of the relative MSE of all models. First, as for the monthly PEEIs, based on the mean of the distribution it is difficult for any linear model to beat the benchmark AR(4) specification in levels with a constant and no trend. Second, and again in line with the monthly results, among the nonlinear and time-varying models the Neural Networks (AN) are the worst. However, now time-varying AR models (ARTV) performs well based on the mean of the empirical distribution for short horizons, generating gains of about 7% for $h = 1$ and 20% for $h = 2$, but losses of at least 25% for h=4. Third, and perhaps most interesting from our perspective, there seem to be some gains from forecast pooling at any forecast horizon. In particular, putting an equal weight on the linear models only yields gains of about 6% for $h = 1$, 13% for $h = 2$, and 8% for $h = 4$. Forecasting with the median from all the models (or from the linear specifications only) is only slightly worse for h<4, but the median is equivalent to the benchmark for h=4. Fourth, if we now rank the models based on the median of the relative MSE across variables rather than the mean, it turns out that for the linear models there are some sizable gains for h=4 from the inclusion of a linear trend in the model, up to 40%. These gains are hidden when looking at the mean since for about 10% of the quarterly PEEIs the linear models with trend perform very badly. Similarly, and in line with the monthly results, for the nonlinear and time-varying models the median is systematically smaller than the mean, since these models perform very poorly for some variables and this deteriorates their average performance. Finally, looking again at the median but this time focusing on the combined forecasts, we find in general comparable results, and this indicates that the fraction of variables for which forecast combination works badly is fairly small.

In summary, the main message is that on average over all the 51 quarterly PEEIs, pooling the linear models using equal weights can yield some gains for any forecast horizon. Interestingly, ARTV models also work well for $h = 1$ and 2. As for the monthly indicators, for specific quarterly PEEIs, the nonlinear or time-varying models can be either very good or very bad, while the performance of pooling appears to be more stable across variables, and this robustness feature can be relevant in an institutional contest.

Additional evidence on the robustness of forecast combination is provided by the fraction of variables for which a given model is in the set of the N models with lowest MSE, for different values of N. If we look, for example, at $h = 1$ and $N = 1$, i.e., the best models for one month ahead forecasts of each indicator, it turns out that the preferred forecast combination on average, the equal weight on the linear models only, is never the best, however it is among the best 20 models for 61% of the quarterly PEEIs. Therefore, forecast combination is not necessarily the best forecasting tool for each indicator, but it is systematically good. On the contrary, complex specific models, such as specific neural networks, can perform very well for a few indicators, but very badly for most of the other indicators. A notable exception is a time varying AR model with constant, three lags, and pre-testing for the unit root, but we should consider that the evaluation sample is fairly short and that similar specifications were not so robust for the monthly indicators, while pooling was quite good also in that case.

As a further check of the robustness of the combination forecasts, we have computed the average loss function in (12.39) for different values of ρ (1, 1.5, 2, 2.5 or 3), where we remind that $\rho = 1$ corresponds to MAE and $\rho = 2$ to MSE. Ranking all the 74 forecasts based on these five loss functions, it turns out that the equal weight combination of linear models is best for $h = 4$, and ranked fourth or fifth for, respectively, $h = 2$ and $h = 1$. The best model for $h = 1$ and $h = 2$ is ARTVFCP3, the time varying AR with constant, three lags, and pre-testing for the unit root mentioned in the previous paragraph, but the same warnings as above apply.

Focusing on quarterly real GDP, a key PEEI, and for $h = 1$, pooling nonlinear forecasts using equal weights minimizes the MSE and performs very well also in terms of MAE and cubed error. Linear specifications are a close second best, and pooling them works reasonably well, with a loss of about 10% in terms of MSE and 5% for MAE with respect to the first best. For $h = 2$ linear specifications become the first best, but pooling nonlinear forecasts is only about 10% worse in terms of MSE and 5% for MAE. Instead, for $h = 4$, a smooth transition model appears to be the first best in terms of both MSE and MAE, with sizable gains over competing specifications.

To conclude, we should stress that the results for quarterly variables should be interpreted with care since the evaluation sample is short, which is also the reason for not conducting stability tests in this case.

12.4 Conclusions

In the first section, we have presented a detailed survey of the literature on forecast pooling, and in the second section we have applied the most promising pooling methods to combine a large set of forecasts for many monthly and quarterly PEEIs.

The main result is that forecast combination provides a good and robust forecast tool for both monthly and quarterly PEEIs. While it is possible to select better forecasting models for each indicator, and these detailed results are provided in the accompanying excel files , it is not possible to find a method that systematically outperforms pooling on average across PEEIs and at any forecast horizon.

Among the pooling procedures, equal weighted combination of linear models works particularly well for the quarterly PEEIs, while assigning higher weight to linear models with lower MSE is even better for the monthly PEEIs.

Bibliography

Aiolfi, M. and Favero, C. A. (2005), Model uncertainty, thick modeling and the predictability of stock returns, Journal of Forecasting, 24, pp. 233-254.

Aiolfi, M. and Timmermann, A. (2004), Persistence of forecasting performance and combination strategies, forthcoming in Journal of Econometrics.

Andrews, D.W.K. and Ploeberger, W. (1994), Optimal tests when a nuisance parameter is present only under the alternative, Econometrica, 62, 1383-1414

Artis, M. and Marcellino, M. (2001), Fiscal forecasting: the track record of IMF, OECD and EC, Econometrics Journal 4:s20-s36.

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

Breusch, T.S. and Pagan, A.R. (1979), A simple test for heteroschedasticity and random coefficient variation, Econometrica, 47, 1287-1294

Chan, Y.L., Stock, J.S. and Watson, M.W. (1999), A Dynamic Factor Model Framework for Forecast Combination, Spanish Economic Review, 1, 1999, pp. 91-121.

Chong, Y.Y. and Hendry, D.F. (1986), Econometric evaluation of linear macroeconomic models, Review of economic Studies, 53, pp. 671-690.

Christoffersen, P. and Diebold, F.X. (1997), Optimal prediction under asymmetrical loss, Econometric Theory, 13, pp. 806-817.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography (with discussion). International Journal of Forecasting 5 , 559-583.

Clemen, R.T., Murphy, A.H. and Winkler, R.L. (1995), Screening probability forecasts: Contrasts between choosing and combining, International Journal of Forecasting, 11, pp.133-145.

Clemen, R.T. and Winkler, R.L. (1986), Combining economic forecasts, Journal of Business and Economic Statistics, 4, pp. 39-46.

Clements, M.P. and Hendry, D.F. (1996), Multi-step estimation for forecasting, Oxford Bulletin of Economics and Statistics, 58:657-684.

Clements, M.P. and Hendry, D.F. (1999), Forecasting non-stationary economic time series, MIT Press, Cambridge, MA.

Clements, M.P. and Hendry, D.F. (1998), Marshall lectures on economic forecasting, Cambridge University Press, Ch. 10.

Coulson, N.F. and Robins, R.P. (1993), Forecast combination in a dynamic setting, Journal of Forecasting, 12, pp. 63-68.

Deutsch, M., Granger, C.W.J. and Terasvirta, T. (1994), The combination of forecasts using changing weights, International Journal of Forecasting, 10, pp. 47-57.

Diebold, F.X. (1988), Serial correlation and the combination of forecasts, Journal of Business and Economic Statistics, 6, pp.105-111.

Diebold, F.X. and Lopez, J. A. (1996), Forecast evaluation and combination, in Maddala and Rao (eds.), Handbook of Statistics (Elsevier, Amsterdam).

Diebold, F.X. and Pauly, P. (1987), Structural change and the combination of forecasts, Journal of Forecasting, 15, pp. 49-61.

Diebold, F.X. and Pauly, P. (1990), The use of prior information in forecast combination, International Journal of Forecasting, 6, pp. 503.508.

Diebold, F.X. and Kilian, L. (2000), Unit Root Tests are Useful for Selecting Forecasting Models, Journal of Business and Economic Statistics, 18:265-273.

Donaldson, R.G. and Kamstra, M. (1996), Forecast combining with neural networks, Journal of Forecasting, 15, pp. 49-61.

Elliot, G. (2004), Forecast combination with many forecasts, Mimeo, Department of Economics, University of California, San Diego.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996), Efficient tests for an autoregressive unit root, Econometrica, 64:813-36.

Elliot. G. and Timmerman. A. (2004), Optimal forecast combinations under general loss functions and forecasts error distributions, Journal of Econometrics, 122, pp. 47-79.

Ghysels, E., Granger, C.W.J. and Siklos P.L. (1996), Is seasonal adjustment a linear or nonlinear data-filtering process?, Journal of Business and Economic Statistics 14:374-386.

Granger, C.W.J. and Ramanathan, M.H. (1984), Improved methods of combining forecasts, Journal of Forecasting, 3, pp. 197-204

Granger, C.W.J. and Newbold, P. (1986), Forecasting economic time series, San Diego: Academic Press.

Granger, C.W.J. and Terasvirta, T. (1993), Modelling nonlinear economic relationships, Oxford: Oxford University Press.

Granger, C.W.J. and Jeon, Y. (2004), Thick modeling, Economic Modelling, 21, pp. 323-343.

Gupta, S. and Wilton, P.C. (1987), Combination of forecasts: An extension, Management Science, 33, pp. 356-372.

Hendry, D.F. and Clements, M.P. (2002), Pooling of forecasts, Econometrics Journal, 5, pp. 1-26.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999), Bayesian model averaging: A tutorial, Statistical Science, 14, pp. 382-417.

Hornik, K., Stinchcombe, M. and White, H. (1989), Multilayer feedforward networks are universal approximators, Neural Networks, 2:359-66.

Koop, G. and Potter, S. (2004), Forecasting in dynamic factor models using Bayesian moving averaging, Econometrics Journal, 7, pp. 550-565.

Leamer, E. (1978), Specification Searches, Wiley, Oxford.

Lee, H.S. and Siklos, P. (1997), The role of seasonality in economic time series-reinterprating money-ouput causality in US data, International Journal of Forecasting 13:381-391.

Makridakis, S. Anderson, A., Carbonne, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. (1982), The accuracy of extrapolation (time series) methods: Results of a forecasting competition, Journal of Forecasting, 1:111-153.

Marcellino, M. (2000) Forecast Bias and MSFE encompassing, Oxford Bulletin of Economics and Statistics, 62, 533-542.

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

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

Marcellino, M., Stock, J.H. and Watson, M.W. (2006), A Comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, Journal of Econometrics,

Meese, R. and Geweke, J. (1984), A comparison of autoregressive univariate forecasting procedures for macroeconomic time series, Journal of Business and Economic Statistics, 2:191-200.

Min, C.K. and Zellner, A. (1993), Bayesian and Non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates, Journal of Econometrics, 56, pp. 89-118.

Newbold, P. and Granger, C.W.J. (1974), Experience with forecasting univariate time series and the combination of forecasts, Journal of Royal Statistical Society A, 137, pp. 131-146.

Nyblom, J. (1989), Testing for constancy of parameters over time, Journal of the American Statistical Association, 84:223-230.

Palm, F.C. and Zellner, A. (1992), To combine or not to combine? Issues of combining forecasts, Journal of Forecasting, 11, pp. 687-701.

Ploeberger, W. and Kramer, W. (1992), The CUSUM test with OLS residuals, Econometrica, 60, 271-286.

Quandt, R.E. (1960), Tests of the hypothesis that a linear regression system obeys two separate regimes, Journal of the American Statistical Association, 55, 324-330.

Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997), Bayesian model averaging for linear regression models, Journal of the American Statistical Association, 92, pp. 179-191.

Stock, J.H. (1996), VAR, error correction and pretest forecasts at long horizons, Oxford Bulletin of Economics and Statistics, 58:685-701.

Stock, J.H. and Watson, M.W. (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in Engle, R. and White, R. (eds), Cointegration, causality, and forecasting: A festschrift in honor of Clive W.J. Granger, Oxford: Oxford University Press, pp. 1-44.

Stock, J.H. and Watson, M.W. (2002), Macroeconomic Forecasting Using Diffusion Indexes, Journal of Business and Economic Statistics, April 2002, Vol. 20 No. 2, 147-162.

Stock, J.H. and Watson, M.W. (2004), Combination forecasts of output growth in a seven-country data set, Journal of Forecasting, 23, pp. 405-430.

Swanson, N.R. and White, H. (1997), A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, Review of Economics and Statistics, 79:540-550.

Terasvirta, T. (1998), Modelling economic relationships with smooth transition regressions in A. Ullah and D.E.A. Giles, eds., Handbook of Applied Economic Statistics (Marcel Dekker, New York) 507-552.

Tiao, G.C. and Xu, D. (1993), Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case, Biometrika, 80:623-641.

Timmermann, A. (2006), Forecast combinations, Handbook of Economic Forecasting, Ch. 4, Vol. 1 (Elsevier)

Winkler, R. L., and Makridakis, S. (1983), The Combination of Forecasts, Journal of the Royal Statistical Society, Ser. A, 146, 150-15

13 Combining Forecasting Techniques: Some Empirical Considerations

Authors

Françoise Charpin
L'institut d'études politiques de Paris (Sciences Po)
francoise.charpin@sciences-po.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

13.1 Introduction

When multiple forecasts of the same variable are available at a same date and over a sufficiently long past, the problem of combining these forecasts should be stated in order to obtain a combination that outperforms the individual forecasts. Unfortunately, in the area of macroeconomic forecasts, that ideal situation practically never occurs. If forecast organisms make regularly forecasts and that since many years, they do not publish them at the same date which makes their forecasts incomparable. There is then a great difficulty to investigate empirically the combined forecast problem by lack of real time forecasts. To overcome this difficulty, we can create our own forecasts. However to be really credible, the forecasts must be carried out in real time. However, some international organizations start to have a complete dataset of real time series, as it is the case for Eurostat, which has a non negligible number of real time series for the euro area. This situation is relatively exceptional given what happens in each euro area country where this kind of data is not available in practice. Nevertheless this information remains limited and it will be difficult to obtain a large variety of forecasts for the same variable. Let us consider for example the GDP and the IPI[1] growths, the two variables chosen hereafter. Most of the forecast studies of these variables will use practically the same set of series, so the information used will only differ marginally. It will then be difficult to obtain multiple non redundant forecasts. In these conditions the diversification gains – which are the principal motivation for forecast combination – may be lacking.

Thus, diversification may be the result of the methods implemented to generate forecasts, hereafter: bridge models, factor models, time series models. Notice that these methods use an increasing information set: a time series model uses no (exogeneous) information, a bridge model is constructed with a small number of individual series and a factor model sums up a larger information set. In this case, it could be expected that factor models outperform other models. Nevertheless, this point should be verified later in this chapter.

It then remains the question of what are the best weights for combining forecasts. On this point, many articles mention that a simple rule of combination may outperform a more complicated one. It probably results from the fact that a weighted mean is not so sensitive to the choice of its weights. Therefore several combination schemes will be evaluated.

The rest of the chapter is organised in two parts: the first one presents an overview about the theory of the "combined forecast" theory, while the second part exposes the practice of combined real time forecast using quarterly GDP growth rate and IPI growth rate data.

13.2 Theory of Forecast Combination

When a number of competing forecasts are available, instead of trying to evaluate the prediction performance of each in order to select one, experience shows that it is preferable to combine them. Each one may contain valuable information not captured by the others, which explains why aggregating them may be of interest. The simpler way to combine them is merely to take the arithmetic mean. However, this procedure is optimal only if all forecasts are unbiased, independent, and identically distributed, which is clearly a very special case. The theory of forecast combination deals with optimal ways of combining forecasts. Hereafter, we only discuss the question of linearly combining linear forecasts. The theory was developed by Bates and Granger (1969) and extended by many others[2], in particular by Granger and Ramanathan (1984), Diebold (1988) and Chan, Stock and Watson (1999). This is still a topical subject as shown by Hendry and Clements (2002).

13.2.1 Combining two unbiased Forecasts

Let $y_t, t \in Z$, be a stationary univariate time series[3] with unconditional mean denoted μ. The case of non-stationary variables is rarely considered in the literature and does not concern us. For simplicity, the majority

[1] Industrial Production Index excluding construction.
[2] As shown in Clemen (1989) and in Timmermann (2006).
[3] We make no notational distinction between random variables and their realisation, using small letters for both.

of papers focus on the case of combining two one-step-ahead forecasts and we also start with two, before generalising to k. The case of combining two h-step-ahead forecasts would be treated similarly. Let $\widehat{y}^1_{t+1,t}$ and $\widehat{y}^2_{t+1,t}$ be alternative forecasts of a given economic time series y_{t+1}, made at time t and with respective errors

$$e^j_{t+1,t} = y_{t+1} - \widehat{y}^j_{t+1,t}, \quad j = 1, 2$$

Suppose first that one has two unbiased forecasts from which a composite is formed as follows.

$$\widehat{y}^c_{t+1,t} = \alpha \widehat{y}^1_{t+1,t} + (1-\alpha)\widehat{y}^2_{t+1,t}$$

Because the weights sum to one, the composite forecast will necessarily be unbiased. The combined forecast error will satisfy the same relation, i.e.

$$e^c_{t+1,t} = \alpha e^1_{t+1,t} + (1-\alpha)e^2_{t+1,t}$$

The error variance is

$$\sigma_c^2 = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2 + 2\alpha(1-\alpha)\rho\sigma_1\sigma_2 \tag{13.1}$$

where

$$Var(e^j_{t+1,t}) = \sigma_j^2, \quad Cov(e^1_{t+1,t}, e^2_{t+1,t}) = \sigma_{12} = \rho\sigma_1\sigma_2$$

The weight α is chosen to minimise the variance σ_c^2, which leads to

$$\alpha = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{13.2}$$

The minimum achievable error variance is

$$\sigma_c^{2*} = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{13.3}$$

Note that $\sigma_c^{2*} < min(\sigma_1^2, \sigma_2^2)$ unless either ρ is exactly equal to $\frac{\sigma_1}{\sigma_2}$ or to $\frac{\sigma_2}{\sigma_1}$. If either equality holds, then the variance of the combined forecast is equal to the smaller of the two error variances. Thus, a priori, it is reasonable to expect in most practical situations that the best available combined forecast will outperform individual forecasts; in any case, it cannot do worse. The benefit of combining is particularly high when the forecast errors are negatively correlated.

Consider now the sign of the weights. Suppose that $\sigma_1 < \sigma_2$, then $0 < \alpha < 1$ if $\rho < \frac{\sigma_1}{\sigma_2}$, $\alpha = 1$ if $\rho = \frac{\sigma_1}{\sigma_2}$ and $\alpha > 1$ if $\frac{\sigma_1}{\sigma_2} < \rho \leq 1$. On the contrary, if $\sigma_2 < \sigma_1$ then $0 < \alpha < 1$ if $\rho < \frac{\sigma_2}{\sigma_1}$, $\alpha = 0$ if $\rho = \frac{\sigma_2}{\sigma_1}$ and $\alpha < 0$ if $\frac{\sigma_2}{\sigma_1} < \rho \leq 1$.

This means that the weights of the combination are positive if the correlation between the forecast errors does not reach a positive bound, which is equal to $\bar{\rho} = \frac{min(\sigma_1,\sigma_2)}{max(\sigma_1,\sigma_2)}$. Otherwise, one weight is negative and the other one is greater than one.

If the weights are constrained to lie between zero and unity, the variance of the combined forecast is equal to the smaller of the two error variances for $\bar{\rho} \leq \rho \leq 1$. In other words, if the forecast errors are too similar (high positive correlation), one gains nothing in combining forecasts, and one retains the forecast with the smallest error variance. In their empirical work, Bates and Granger (1969) and Granger and Newbold (1977)

constrain the weights to be non-negative, pointing out that it is rather difficult to justify on general grounds the assignment of a negative weight to a particular forecast. Bates and Granger take the following example: if two forecasts are eighty and one hundred and the respective weights are minus one and two, the combined forecast is one hundred and twenty, greater than the components.

However, Granger and Newbold do not reject the case of a negative weight on the basis of the following example. They first note that when ρ approaches 1 or 1, a perfect forecast is obtainable and they examine the case $\rho = 1$ (the other case leading to positive weights). They consider two forecasts producing errors e and $2e$ for example. Then

$$e = y_{t+1} - \widehat{y}^1_{t+1,t} \text{ and } 2e = y_{t+1} - \widehat{y}^2_{t+1,t}$$

Eliminating e between the two equations leads to $y_{t+1} = 2\widehat{y}^1_{t+1,t} - \widehat{y}^2_{t+1,t}$

In this textbook case, the weights 2 and -1 are then optimal. It is clear that when the forecast errors are highly positively correlated, attributing a negative weight to the inferior forecast permits a reduction in variance through the covariance because it is then preceded by a minus sign in equation (13.1). In my opinion, this is an artificial way to reduce the error variance. However, this point of view is not shared, because in the majority of the papers the question of the sign of the weights is not even evoked.

For empirical implementation, the weight α is estimated by replacing variances and covariance in (13.2) with their sample estimates

$$\widehat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} e^i_{t+1,t} e^j_{t+1,t} \text{ with } \widehat{\sigma}_{ii} = \widehat{\sigma}^2_i \tag{13.4}$$

yielding to

$$\widehat{\alpha} = \frac{\widehat{\sigma}^2_2 - \widehat{\sigma}_{12}}{\sigma^2_1 + \sigma^2_2 - 2\widehat{\sigma}_{12}}$$

In small samples (which is often the case when combining forecasts), sampling error contaminates the weight estimate, and the problem of sampling error is exacerbated by the collinearity that often exists among primary forecasts. Thus, while one hopes to reduce out-of-sample forecast mean square error (MSE) through combination, there is no guarantee of it. In practice, it turns out that forecast combination techniques often perform well, as documented in Clemen's (1989) review. However, the uncertainty about the sample estimate leads to the consideration of a number of alternative choices of weights.

For example, Bates and Granger (1969) find that considering $\widehat{\alpha} = \frac{\widehat{\sigma}^2_2}{\sigma^2_1 + \sigma^2_2}$, i.e. ignoring the correlation between the forecast errors ($\sigma_{12} = 0$), is considerably more successful than attempting to take it into account.

13.2.2 Combining k unbiased forecasts

The results discussed earlier can be extended to the combination of more than two forecasts. A single series y_{t+1} is to be forecasted using combinations of k forecasts, $\widehat{y}^j_{t+1,t}, j = 1, 2, \ldots, k$, with $t = 1, \ldots, T$. Assume as before that these forecasts are unbiased. The notations used are

$$\widehat{\mathbf{y}}'_{\mathbf{t+1,t}} = (\widehat{y}^1_{t+1,t}, \widehat{y}^2_{t+1,t}, \ldots, \widehat{y}^k_{t+1,t})$$

$$\mathbf{u}' = (1, 1, ..., 1)$$

$$\mathbf{e_{t+1,t}} = y_{t+1}\mathbf{u} - \widehat{\mathbf{y}}_{\mathbf{t+1,t}} \text{ with } E(\mathbf{e_{t+1,t}}\mathbf{e}'_{\mathbf{t+1,t}}) = \Sigma$$

The combined forecast can be written as a weighted average of individual forecasts

$$y^c_{t+1,t} = \alpha' \widehat{\mathbf{y}}_{\mathbf{t+1,t}} \text{ with } \alpha' \mathbf{u} = \mathbf{1}$$

The variance of the combined forecast error is:$Var(\alpha' \mathbf{e_{t+1,t}}) = \alpha' \Sigma \alpha$

Without sign constraint on the weights

When the variance is minimised under the constraint $\alpha' \mathbf{u} = 1$ and without any sign constraint on the weights α_j, the optimal weights are given by:

$$\alpha = \frac{1}{\mathbf{u}' \sum^{-1} \mathbf{u}} \Sigma^{-1} \mathbf{u} \tag{13.5}$$

and the minimum variance of the combined forecast error, by $(\mathbf{u}' \Sigma^{-1} \mathbf{u})^{-1}$.

Constraining the weights to be non negative

When the constraints $\alpha_j \geq 0$, $j = 1, \ldots, k$ are imposed, the optimal solution has a similar expression, but written for the subset of indices for which the α_j are strictly positive at the optimum.

Let $J = \{j | \alpha_j > 0\}$ and $I = \{j | \alpha_j = 0\}$. The vectors and the covariance matrix are partitioned according to the two index subsets and let $\alpha_{\mathbf{J}}$, $\mathbf{u_J}$ and $\mathbf{\Sigma_{JJ}}$ be the blocks of interests. The optimal weights are then given by

$$\alpha_{\mathbf{J}} = \frac{1}{\mathbf{u}'_{\mathbf{J}} (\Sigma_{JJ})^{-1} \mathbf{u_J}} \Sigma^{-1}_{JJ} \mathbf{u_J} \tag{13.6}$$

When k is large, the optimal partition $\{J, I\}$ can be found using quadratic programming software. Since the elements of S will not be known in practice, they have to be estimated by their sample estimates given in (13.4).

Ignoring correlations across forecast errors

In this case,[4] the optimal weights are found to be all positive and expression (13.5) leads to

$$\alpha_j = \frac{\frac{1}{\sigma_j^2}}{\sum_{i=1}^{k} (\frac{1}{\sigma_i^2})}, \quad j = 1, \ldots, k \tag{13.7}$$

Thus, the weights are equal to their relative inverse mean squared errors.

Identical pair-wise correlations

Instead of assuming that all pair-wise correlations are null, it is more realistic to take them equal (and positive) but different from zero. In this case, the optimal weights are not necessarily non-negative and (13.6) must be considered if (13.5) leads to some negative weights.

Identical pair-wise correlations and identical variances

In this unrealistic case the optimal weights are found to be equal: $\alpha_j = 1/k, \quad j = 1, \ldots, k$

[4] Originally introduced by Bates and Granger (1969).

13.2.3 An alternative presentation within the regression framework

When the weights are not constrained to be non-negative, the forecast combination method can be viewed within a regression framework.

The empirical variance of the combined forecast error is given by

$$\frac{1}{T}\sum_{t=1}^{T}(y_{t+1} - \alpha_1 \widehat{y}^1_{t+1,t} - \alpha_2 \widehat{y}^2_{t+1,t} - \ldots - \alpha_k \widehat{y}^k_{t+1,t})^2 \tag{13.8}$$

which can be rewritten using the constraint that the weights sum to one

$$\begin{aligned}\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\big[(y_{t+1} - \widehat{y}^k_{t+1,t}) - \alpha_1(\widehat{y}^1_{t+1,t} - \widehat{y}^k_{t+1,t}) - \alpha_2(\widehat{y}^2_{t+1,t} - \widehat{y}^k_{t+1,t}) \\ - \ldots - \alpha_{k-1}(\widehat{y}^{k-1}_{t+1,t} - \widehat{y}^k_{t+1,t})\big]^2\end{aligned} \tag{13.9}$$

The optimal weights are obtained by minimising (13.9), which corresponds to unrestricted least squares with no intercept. They can also be obtained by minimising (13.8) subject to the constraint $\sum_j \alpha_j = 1$, using restricted least squares with no intercept. As in Nelson (1972), the k weights α_j can first be estimated by unrestricted least squares with no intercept using (13.8), then the validity of the constraint $\sum_j \alpha_j = 1$ can be tested, before using restricted least squares if the constraint is accepted.

The unrestricted approach as a bias correction

In the previous sections, it was supposed that the individual forecasts were unbiased. If this is not the case, it is recommended to unbias them before combining. For that, the following regressions can be run

$$y_{t+1} = a_j + b_j \widehat{y}^j_{t+1,t} + \eta^j_{t+1}$$

and one can check whether $\{a_j = 0, b_j = 1\}$,[5] and whether is white noise. If any of these conditions do not hold, an immediately apparently superior forecast can be achieved and these should be used in any combination. This advice is given by Granger (1989) and it allows one to think that only unbiased forecasts are combined.

Nevertheless Granger and Ramanathan (1984) justify the new method they propose by the fact that, when the component forecasts are biased, the combination is unbiased. They first note that the combining weights were chosen to add to one, on the assumption that each forecast was unbiased, making the combination unbiased. If now the component forecasts are biased, there is no reason to constrain the combining weights to sum to one, and they suggest running unconstrained least squares with an intercept (instead of forcing the regression through the origin), that is

$$y_{t+1} = \alpha_0 + \alpha_1 \widehat{y}^1_{t+1,t} + \alpha_2 \widehat{y}^2_{t+1,t} + ... + \alpha_k \widehat{y}^k_{t+1,t} + \epsilon_{t+1} \tag{13.10}$$

The inclusion of an intercept implies the following unbiased combined forecast

$$\widetilde{y}^c_{t+1,t} = \widetilde{\alpha}_0 + \sum_{j=1}^{k} \widetilde{\alpha}_j \widehat{y}^j_{t+1,t}$$

Moreover, the unconstrained regression will give a better fit and presumably better forecasting performance. They conclude "The common practice of obtaining a weighted average of alternative forecasts should, there-

[5] This is a sufficient condition for unbiased-ness. Unbiased-ness only requires $a_j = (1 - b_j)E(\widehat{y}^j_{t+1,t})$

fore, be abandoned in favour of an unrestricted linear combination including an intercept". Their conclusion has often been contested, see Clemen (1989) page 562. Obviously, to have better in-sample performance in terms of root mean square forecasting error does not guarantee better out-of-sample performance. For example, Clemen (1986) shows that restricting the combination of US GNP forecasts improves the combined forecast performance compared to the unrestricted combination. He suggests, within the unrestricted regression framework, testing the validity of the constraints $\sum_{j=1}^{k} \alpha_j = 1$ and $\alpha_0 = 0$, and to impose both or only the first one, if the individual forecasts are biased.

The Diebold contribution

The difference between the two methods (restricted and unrestricted regression) is well analysed in Diebold (1988) and Diebold and Lopez (1996). For this, it is necessary to return to the forecast errors. The restricted regression defines the combined forecast presented in (13.11)

$$\widehat{y}^{c}_{t+1,t} = \sum_{j=1}^{k} \widehat{\alpha}_j \widehat{y}^{j}_{t+1,t} \text{ with } \sum_{j=1}^{k} \widehat{\alpha}_j = 1 \tag{13.11}$$

and the unrestricted regression presented in (13.12),

$$\widetilde{y}^{c}_{t+1,t} = \widetilde{\alpha}_0 + \sum_{j=1}^{k} \widetilde{\alpha}_j \widehat{y}^{j}_{t+1,t} \tag{13.12}$$

With definition (13.11) the combined and individual forecast errors will satisfy the same relation

$$\widehat{e}^{c}_{t+1,t} = \sum_{j=1}^{k} \widehat{\alpha}_j e^{j}_{t+1,t} \text{ with } \sum_{j=1}^{k} \widehat{\alpha}_j = 1 \tag{13.13}$$

whereas with definition (13.12), the combined forecast error cannot be expressed with solely the individual forecast errors $e^{j}_{t+1,t}$, $j = 1, \ldots, k$. However, it will appear to be a combination of forecast errors as shown by Diebold (1988).

Before examining this point, one can make several remarks. Minimising the variance of (13.13) corresponds to the principle that by diversification one reduces risk. Finding the weights of (13.12) using regression corresponds to the search for the best forecast equation starting from several forecasts instead of explanatory variables. The spirit of the two methods differs, but both are conceivable. Note that if all forecast errors are null, the combined forecast error in (13.13) will also be null, whereas with definition (13.12), this property will not be satisfied. Similarly, if all individual forecasts are equal, the combined forecast will have the same value using (13.11) but not with (13.12).

With definition (13.12), the combined forecast error is

$$\widetilde{e}^{c}_{t+1,t} = y_{t+1} - \widetilde{\alpha}_0 - \sum_{j=1}^{k} \widetilde{\alpha}_j \widehat{y}^{j}_{t+1,t} \; with \; \widetilde{\alpha}_0 = E(y_{t+1}) - \sum_{j=1}^{k} \widetilde{\alpha}_j E(\widehat{y}^{j}_{t+1,t})$$

Let $\bar{e}_{t+1,t}$, $j = 1, \ldots, k$ be the forecast errors corrected for their bias b_j, $(j = 1 \ldots k)$.

$$\bar{e}^{j}_{t+1,t} = e^{j}_{t+1,t} - b_j$$

Then

$$\widetilde{\alpha}_0 = \mu - \sum_{j=1}^{k} \widetilde{\alpha}_j(\mu - b_j) = (1 - \sum_{j=1}^{k} \widetilde{\alpha}_j)\mu + \sum_{j=1}^{k} \widetilde{\alpha}_j b_j$$

and

$$\widetilde{e}^c_{t+1,t} = y_{t+1} - (1 - \sum_{j=1}^{k} \widetilde{\alpha}_j)\mu - \sum_{j=1}^{k} \widetilde{\alpha}_j b_j - \sum_{j=1}^{k} \widetilde{\alpha}_j (y_{t+1} - b_j - \bar{e}^j_{t+1,t})$$

Finally

$$\widetilde{e}^c_{t+1,t} = (1 - \sum_{j=1}^{k} \widetilde{\alpha}_j)(y_{t+1} - \mu) + \sum_{j=1}^{k} \widetilde{\alpha}_j \bar{e}^j_{t+1,t} \tag{13.14}$$

This shows that the combined forecast error may be decomposed into a linear combination, with weights summing to unity, of the forecast errors of the unconditional mean forecast and the bias-corrected primary forecast errors. Thus the unrestricted regression method amounts to the inclusion of one more forecast among those to be combined—the unconditional mean. One can ask if this inclusion is beneficial. Diebold thinks that it is. He gives the following argument “noisy data tend to produce volatile combining weights, so the incorporation of some prior information may provide valuable robustness”. This is the principle of shrinkage.

However, this method has at least one drawback: the residual of the unrestricted regression (i.e., the combined forecast error) will most often present serial correlation. This is precisely the subject of Diebold’s paper. If we look at equation (13.14), we see that the autocorrelation may arise if y is serially correlated which is more than probable. Thus Diebold proposes taking this serial correlation into account when estimating the unrestricted regression, i.e., using generalised least squares with an error term following an AR(p) process with $p = 1$ in his paper. The author presents an example with simulated data where the two primary forecasts are unbiased. He compares the simple arithmetic average combined forecast, the restricted OLS combined forecast, the unrestricted OLS combined forecast and the unrestricted AR combined forecast. It turns out that the simple average and the restricted OLS clearly beat the unrestricted OLS. But the best results are obtained with the unrestricted AR combined forecast. The unrestricted AR combined forecast is nearly equivalent to introducing into the unrestricted regression lagged dependent variables and lagged forecasts, a method chosen by Coulson and Robbins (1993).

13.2.4 Bias correction in the optimisation framework

The unrestricted regression method was first presented by Granger and Ramanathan (1984) as a "bias correction". A bias correction can also be obtained in the optimisation framework presented in Section 13.2.2. It suffices to add the constraint expressing that the combined forecast is unbiased. The optimal vector of weights is then the solution of the optimisation problem

$$Min\ \alpha' \sum \alpha$$

$$\sum_j \alpha_j = 1]$$

$$\sum_j \alpha_j E(\widehat{y}^j_{t+1,t}) = \mu$$

with possibly $\alpha_j \geq 0,\ j = 1, 2, \ldots, k$

The component forecasts can include those given by an ARMA model. The advantage of using optimisation instead of regression is the possibility of constraining the weights to be non-negative, i.e., the possibility of excluding from the combination some superfluous forecasts, in the sense of being too highly correlated with others given their variance. Allowing for the possibility of negative weights corresponds to the wish to exploit

these high positive correlations. But is this a good idea, given that correlation coefficients are generally not well estimated in small samples?

13.2.5 Further Extensions

The suggestion by Granger and Ramanathan (1984) of combining forecasts by simply regressing realisations on forecasts gives rise to multiple variations and extensions because any "regression tool" is potentially applicable. The extensions are time-varying combining weights, dynamic combining regressions, Bayesian shrinkage of combining weights toward equality and non-linear combining regressions. These generalisations are not always convincing. Empirical studies show that simple rules for combining forecasts, such as averages (i.e., equal weights), often work as well as more elaborate rules based on the relative past performance of the forecasts to be combined. These empirical findings are no incitement to examine sophisticated methods.

All methods described until now do not require particular assumptions on forecasts or forecast errors except for the stationarity and, sometimes, the unbiased-ness of individual forecasts. The methods proposed in the next section make the assumption that either forecast errors or forecasts follow a factor structure.

13.2.6 Methods based on factor structures

Figlewski's contribution (Figlewski (1983))

The objective of the author is to put restrictions on the covariance matrix of the forecast errors because this matrix is poorly estimated, by lack of data. The covariance matrix contains $k(k+1)/2$ distinct elements and an accurate estimate will require an enormous amount of data. Moreover the empirical covariance matrix is invertible only if $T > k$. In the case of financial analyst forecasts, k is usually large and not necessarily inferior to the number T of observed forecasts. The problem of parameterising a covariance matrix by minimising the number of parameters to be estimated arises in finance with the covariance matrix of security returns. What is done is to assume a one-factor model for security returns, namely the market model. The same approach can be used for forecasting. Figlewski proposes adopting the equivalent of the market model to explain individual forecast errors as follows

$$e^j_{t,t-1} = \beta_j \delta_t + \epsilon_{jt} \tag{13.15}$$

where $\delta_t \sim N\left(0, \theta^2\right)$, $\epsilon_{jt} \sim N\left(0, s_j^2\right)$, and δ_t, ϵ_{jt} and ϵ_{it} are mutually independent (at all lags and leads).

The common factor δ_t represents an inability to anticipate some events which all forecasters share. It measures the systematic forecast error in each period. The random variable ϵ_{jt} is the specific error made by each forecaster.

With this model, the elements of the error covariance matrix becomes

$$\sigma_{ij} = \beta_i \beta_j \theta^2 \; and \; \sigma_i^2 = s_i^2 + \beta_i^2 \theta^2 \tag{13.16}$$

The number of parameters needed to calculate the optimal combination is greatly reduced, to $(3k+1)$ parameters.

The first step is to estimate the single factor δ. Figlewski proposes taking in each period the mean error $\delta_t = \frac{1}{k}\sum_j e^j_{t,t-1}$. Then, the parameters β_j and the specific errors ϵ_{jt} are obtained by regressing the forecast errors on an intercept and on the factor δ. Finally, the variances and covariances given in (13.16) can be estimated. The optimal weights are chosen following Bates and Granger (1969).

The drawbacks of this work are the assumption of no correlation between specific errors (strict factor model) and the way the common factor is estimated. The model could have been developed with the assumption of

correlations between specific errors (approximate factor model), which would have led to estimate the common factor differently. In fact, instead of introducing some variation into the Figlewski model, it seems preferable to follow Chan, Stock and Watson (1999). They choose to model the forecasts by a factor model instead of modelling the forecast errors.

The Chan, Stock and Watson contribution (1999)

The panel of individual forecasts is modelled as an approximate factor model where the unobserved factor is the true conditional expectation. Let μ_t be the conditional expectation of $(y_{t+1} - \mu)$.[6] The forecasts are assumed to be unconditionally unbiased and are assumed to be decomposed into three terms:

$$\widehat{y}^{j}_{t+1,t} - \mu = \mu_t + \zeta_{jt} + \nu_{jt} \tag{13.17}$$

The expression (13.17) represents the difference between the jth forecast (mean-adjusted) and the optimal forecast as the sum of two components. The first one, ζ_{jt}, represents the model specification error. The second component, ν_{jt} , represents the estimation error. They suppose that, generally, μ_t will be correlated with $\zeta_{jt} + \nu_{jt}$. Because of this correlation, $E(\widehat{y}^{j}_{t+1,t} - \mu \,|\, \mu_t) \neq \mu$. If $(\mu_t, \zeta_{jt} + \nu_{jt})$ are jointly normally distributed, then this conditional expectation is linear and (13.17) implies,

$$\widehat{y}^{j}_{t+1,t} - \mu = \lambda_j \mu_t + \epsilon_{jt} \tag{13.18}$$

where $\lambda_j = 1 + \frac{Cov(\zeta_{jt}+\nu_{jt},\mu_t)}{Var(mu_t)}$ and ϵ_{jt} is normally distributed and is independent of μ_t.

In matrix notation, (13.18) can be written

$$\widehat{\mathbf{y}}_{t+1,t} - \mu\mathbf{u} = \lambda\mu_t + \epsilon_t \tag{13.19}$$

where $\widehat{\mathbf{y}}_{t+1,t}$ denotes the k-vector of forecasts, λ is a k-vector with element λ_j, and $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{kt})$.

Given the structure (13.18), what is the optimal way to choose weight α? The optimality criterion is, as previously, the minimisation of the mean squared forecast error (equal to the variance for unbiased forecasts). It leads to

$$\alpha = [\sum_{\epsilon} + \sigma^2_{\mu}\lambda\lambda^{'}]^{-1}\,\sigma^2_{\mu}\lambda \tag{13.20}$$

where $\sigma^2_{\mu} = E(\mu^2_t)$ and $\sum_{\epsilon} = E(\epsilon_t\epsilon^{'}_t)$

Unfortunately, equation (13.20) cannot be computed because it contains unknown parameters.

So the authors propose exploiting the potential factor structure of the forecasts without considering (13.20), which corresponds to the minimisation of the variance error. The best combined forecast is theoretically the conditional mean μ_t. The objective is then to estimate μ_t, which is possible with the factor structure assumption of the forecasts. The evaluation of the performance of this method will be conducted using Monte Carlo techniques for a particular set of parameters.

Stock and Watson (1998) show that the principal component of the forecast second-moment matrix will be a consistent estimator of the single common factor as the sample size and k get large, and OLS regression with this principal component will produce an asymptotically efficient estimate of μ_t. If b is the eigenvector corresponding to the largest eigenvalue of the covariance matrix of the forecasts, then the first principal

[6] In the paper of Chan, Stock and Watson, it is supposed that the unconditional mean of y_{t+1} is zero. As we have previously supposed that the unconditional mean of y_{t+1} was μ, we replace in their equations y_{t+1} by $(y_{t+1} - \mu)$.

component is $b^{'}(\widehat{y}_{t+1,t} - \bar{\widehat{y}})$, where $\bar{\widehat{y}}$ is the vector of forecast means. But this component is not appropriately scaled to represent an estimate of μ_t . Given the definition of the conditional mean μ_t

$$\widehat{y}_{t+1,t} - \mu = \mu_t + \epsilon_{y,t+1}$$

we infer that the scale problem will be solved by regressing y_{t+1} on the first principal component and an intercept

$$\widehat{y}_{t+1} = c + \gamma b^{'}(\widehat{y}_{t+1,t} - \bar{\widehat{y}}) + error_{t+1} \tag{13.21}$$

The weights α are given by $\alpha = \gamma b$. The combined forecast, i.e. the estimate of μ_t , can be written

$$\widehat{y}^{c}_{t+1} = \bar{y} + \gamma(\sum_{j=1}^{k} b_j(\widehat{y}^{j}_{t+1,t} - \bar{\widehat{y}}^{j})) \tag{13.22}$$

The combined forecast is unbiased. The weights $\alpha_j = \gamma b_j$ do not necessarily sum to one. All forecasts have strictly positive weights, insofar as they are positively correlated with the variable y.

The objectives of the Monte Carlo analysis are to evaluate various combination forecasting methods when the forecasts are taken to follow the factor model (13.19). Six combination forecasting methods are considered: equal weighting, unrestricted OLS, the James-Stein estimator; ridge regression; principal component regression, and the median. The conclusion is that the PC forecast is most often the best forecast and is always close to the best forecast. The equal-weighting estimator generally works well unless k is small. The median works well for $k \geq 10$, and is the best of those considered when there are some large outliers in the forecasts.

After the simulation study, the authors present an empirical investigation of combination forecasts using a dataset of simulated real-time univariate forecasts of 215 monthly macroeconomic time series for the United States. They find that the PC and the OLS forecasts perform poorly. The simple average forecasts work especially well and the median perform well when the dataset contains some large outliers.

These last results are not very encouraging and provide us with no incitement to adopt either the approaches based on the minimisation of the error variance, or those based on a factor structure of the forecasts.

13.2.7 Forecast encompassing test

Combining forecasts is not necessarily interesting, for example, when you find one optimal weight close to one and the other ones close to zero. In this case, a forecast model dominates all other ones. Chong and Hendry (1986) propose a test to compare two models and to conclude if one model is better than the other, such that combining them has no interest. The test is whether or not the forecast errors of a given model (h) may be explained (at least in part) by the forecasts of another model (l). It is the test for $b = 0$ (with the t-ratio and the $N(0, 1)$ distribution) in the following regression:

$$y_{t+1} - \widehat{y}^{h}_{t+1,t} = b\widehat{y}^{l}_{t+1,t} + \epsilon_{t+1} \tag{13.23}$$

If the null hypothesis is accepted, model h's forecast errors are not predictable by model l's forecasts, then model h "forecast-encompasses" model l. That is, given model h's forecasts, model l's forecasts are redundant for predicting the variable of interest.

Ericsson and Marquez (1993) generalizes the Chong and Hendry's test, allowing for

(a) an intercept in (13.23)

(b) comparison against several models at once, rather than just one

(c) model nonlinearity

(d) multi-step ahead forecasts from dynamic models

(e) the uncertainty from estimating, rather than knowing, model coefficients.

The regression accounting for (a) and (b) is

$$y_{t+1} - \widehat{y}^{h}_{t+1,t} = b_0 + \sum_{k \neq h} b_k \widehat{y}^{k}_{t+1,t} + \epsilon_{t+1} \tag{13.24}$$

By allowing for a nonzero intercept, b_0, and testing the nullity of the b_k coefficients, a more powerful forecast-encompassing test may result if the forecast errors of model h are systematically biased. It is also of interest to test hypotheses $b_0 = 0$ and $b_k = 0$ jointly.

Ericsson (1992) shows that having the smallest RMSFE across a set of models is a necessary (but not sufficient) condition for forecast-encompassing those models. That is, forecast encompassing is a more stringent criterion than having the smallest RMSFE.

There is an obvious link between forecast encompassing theory and forecast combination. If we compare (13.24) to (13.10), the forecast encompassing test corresponds to the test of the null hypothesis:

$$\{\alpha_h = 1 \text{ and } \alpha_k = 0, \forall k \neq h\}$$

as stated in equation 13.10.

As we mentioned in the beginning of this section: combining forecasts is not interesting when you find one optimal weight close to one and the other ones close to zero. In this case, model h dominates all other models.

13.3 Combination of real time forecasts in practice

13.3.1 The cases of the euro area GDP and IPI

The problem in evaluating the performance of forecast combination methods comes from the lack of available forecasts. Thus, the first step is to generate forecasts with several structural[7] models, and the second, is to study the benefits from forecast combination. Simulated forecasts from the structural models are carried out with real time data to obtain the forecasts as they would have been generated in the past. We carry out these real time forecasts for two euro area series: the quarterly GDP growth rate and the monthly IPI growth rate. The models are constructed using the real time data available from Eurostat (European Commission) and related to the PEEIs[8] dataset. Five models[9] concern the quarterly GDP growth rate, and three[10], the monthly IPI growth rate. All give forecasts over the 2002-2007 period. The quarterly models provide one-step-ahead forecasts and the monthly, one and two-step-ahead forecasts.

This part is structured as follows. Section 13.3.2 (respectively Section 13.3.4) describes the forecasting models under investigation for the quarterly GDP growth rate (respectively for the the monthly IPI growth rate). Section 13.3.3 (respectively Section 13.3.5 for IPI) presents an evaluation of forecasts (absence of bias, root mean squared forecast errors, encompassing tests) and compares some combinations of these forecasts.

13.3.2 Forecasting models of the quarterly GDP growth rate

Four models are regressions, denoted from M1 to M4, the fifth model M5 is the usual benchmark $AR(1)$ model[11]. Two regression models use individual series as regressors (M1 and M2). The two others use factors (i.e. principal components) as regressors (M3 and M4). At present, Eurostat releases a flash estimate of GDP for quarter T, published around the middle of the second month of quarter $(T+1)$. We produce here a first estimate for quarter T at the end of the second month of quarter T (denoted d_1), a second more reliable estimate at the end of the third month of quarter T (denoted d_2) and a third estimate, at the end of the first month of quarter $(T+1)$ (denoted d_3). So each indicator $(j = 1, ..., 5)$ gives successively three estimates of GDP for quarter T, denoted $f^{(j)}_{T,d_1}$, $f^{(j)}_{T,d_2}$ and $f^{(j)}_{T,d_3}$, respectively at the dates d_1, d_2 and d_3.

Bridge Models

The first model M1 includes the industrial production[12] growth rate as regressor, contrary to the second model M2. The second model is built to answer the question: Can we estimate GDP without using IPI? In theory introducing IPI is a good option because IPI is a component of GDP and is used by many euro area countries to produce their flash GDP estimates. But in practice, this generates two difficulties. Industrial production is subject to rather long publication delays (industrial production for month $(m-2)$ is released at mid-month (m) and to substantial revisions. The delay implies that it is necessary to forecast IPI[13] and the revisions imply some variability in resulting GDP estimates. In the second model, IPI is replaced by the industrial confidence index because it is the main series relevant to forecast IPI. Apart these coincident series, the two models include the same leading regressors (see Table 13.1) namely the confidence index of the construction survey, the household's opinion on major purchases and only one leading financial series, the real dollar-euro exchange rate.

[7] The term "structural" is excessive for our types of model, but this is convenient to use it in order to make the distinction with time series models that have no economic contain.
[8] Principal European Economic Indicators
[9] Four structural models and one time-series model.
[10] Two structural models and one time-series model.
[11] Autoregressive terms of order greater than one are not significant.
[12] This is the "industrial production excluding construction" series.
[13] Unfortunately, IPI forecasts are not very accurate due to the high volatility of the series.

Table 13.1: Coincident and leading series used in the two bridge models M1 and M2

Regressors	Lag
M1 : Industrial production index[a] (growth rate)	0
M2 : Change in industrial confidence index	
Change in households' opinion on major purchases over next 12 months	1
Change in construction confidence index	3 and 4
Real dollar/euro exchange rate (growth rate)	2

[a]Excluding construction

On the day of producing a coincident GDP indicator for quarter T, industrial production data covers possibly two (at d_3), one (at d_2) or zero months (at d_1) of this quarter. It is thus necessary to forecast industrial production for the missing months, which will be done with the regression model (P1) described in Section 13.3.4. On the other hand, when survey data are used, one month of surveys is missing for the first GDP estimation (at d_1). For the other two estimation dates, survey data are completely known for the quarter to be estimated. When one month of survey data is missing, we use the average of the two available months to replace the average of the three.

We now describe the out-of-sample estimation errors done with these two models using real-time data, over the period (2002-2007). For example, the GDP of the first quarter 2002 is obtained with data available at the end of year 2001 and at the beginning of 2002 (end of January), etc., until the GDP of the fourth quarter 2007, obtained with data available at the end of year 2007 and at the beginning of 2008 (end of January). Thus 72 regressions are run per model. All estimation errors are computed with the GDP flash estimate growth rates.

Factor Models

Our factor models M3 and M4 are both constructed in the spirit of the Stock and Watson (2002). Namely, a data set is re-organized into principal components. These factors are introduced in a regression to explain GDP growth (possibly with lags) and those which are statistically significant are kept. The main difference between our models and the usual ones (inspired from the Stock and Watson's article) is that our data set is small. Usually principal components are extracted from a large data set. But it seems that the introduction of many series, more or less related to GDP, can produce a noise that deteriorates the estimate – see Boivin and Ng (2006), Bai and Ng (2008). Hence we propose to consider only the series directly related to GDP growth, approximately all these which can help to predict the GDP growth but which cannot be introduced simultaneously in a regression because of multicollinearity. These targeted predictors are chosen according to two ways that gives rise to our two models M3 and M4.

For the model M3, Table 13.2 presents the nine selected series. Moreover these series are lagged if they entered in our regression models with a lag. As they cannot be simultaneously used in a regression because of their collinearity, we extract the principal components (PC's) of the data set. This is a way of keeping all these individual series directly related to GDP growth rate. The extraction of PC are carried out on standardized data, i.e. that we compute the eigen vectors and values of the correlation matrix. Then we regress the GDP growth rate on these nine PC and an intercept. We finally select the PCs which are significant. This model is developed in Charpin, Mathieu and Mazzi (2008).

For the model M4, following Bai and Ng (2008) [14], the targeted predictors are selected using the LARS algorithm (Least Angle Regression). This algorithm can be considered as the successor of stepwise algorithms (forward selection, forward stepwise regression). These usual algorithms tend to be too aggressive in the sense of eliminating too many predictors correlated with the ones included. The LARS algorithm has not this drawback and it allows us to keep correlated series which is desirable when the final objective is to extract PC. The LARS algorithm is presented in Efron, Hastie, Johnstone and Tibshirani (2004). A short description is also given by Bai and Ng (2008). The LARS algorithm will give us a ranking of the potential predictors when the presence of other predictors is taken into account.

[14]The term "targeted predictors" is introduced by these authors.

The first difference between our implementation of the LARS algorithm and the one of Bai and Ng (2008) is that some series are introduced several times in our data set, but with different lags, contrary to the Bai and Ng's data set where only lag 0 is considered for all series. Indeed, we expect the LARS algorithm to reveal the coincident or leading character of a series. The second difference which concerns only the survey data model, a series is introduced both in level[15] and in variation in the data set. The LARS algorithm helps in choosing the form of a series (level or change or both). The combination of several lags and several forms multiply the number of variables in the data set. One advantage of the LARS algorithm is its ability to class very rapidly a great number of series starting from the most predictive to the less predictive according to the selection criterion of this algorithm. The third difference between our work and the one of Bai and Ng (2008) is that the principal components are extracted from a data set containing no financial series. These latters are introduced directly in the GDP regression beside principal components as they would generate their own distinct PC. In this case it is preferable to consider directly the financial series instead of its associated PC. Finally, the PC are extracted from a small data set containing the ten series presented in Table 13.3. As previously, the PC extraction is carried out on standardized data. We then regress the GDP growth rate on these ten PC's and an intercept. We finally select the significant PC's. We add to these factors only one financial data, the growth rate of the real dollar/euro exchange rate (with two quarters lagged). As in the model M3, it is the unique financial data which helps to predict GDP growth.

As previously, we have to forecast industrial production for the missing months (Model P2 in Section 13.3.4). When one month of surveys is missing, we use the average of the two available months to replace the average of the three.

Table 13.2: Coincident and leading series used to construct PC of the factor model M3

Series	Lag
Industrial production index (exc. construction) (growth rate)	0
Change in industrial confidence index	0
Households' financial situation over next 12 months	0
Change in households' opinion on major purchases over next 12 months	1
Change in construction confidence index)	3 and 4
Change in employment expectations in construction	3 and 4
Real dollar/euro exchange rate (growth rate)	2

Table 13.3: Coincident and leading series used to construct PC of the factor model M4 according to their LARS rank at the end of 20

Series	Lag
Industrial production index (exc. construction) (growth rate)	0
Economic sentiment indicator	0
General economic situation over the next 12 months (consumer survey)	0
Change in opinion on assessment of order books (construction survey)	2
Change in households' opinion on major purchases over next 12 months	1
Expected business situation (retail survey)	0
Change in opinion on present business situation (retail survey)	0
Consumer confidence indicator	0
Change in households' opinion on financial situation over next 12 months	0
Change in employment expectations (construction survey)	0

The out-of-sample estimation errors done with these two factor models using real-time data are computed over the 2002-2007 period. Thus, we perform 72 principal component analyses and 72 regressions. For M3, all regressions include the first three factors[16], none of them includes the sixth, eighth and ninth factors[17]. On average, four or five PC's are present in the 72 regressions. Let us note that this method does not give better fits than a regression with individual series. However, for out-of-sample estimations, this could be better even if the in-sample estimation is not. Its potential superiority derives from being estimation less dependent on

[15] In this case, the level must be stationary.

[16] The PC's are ranked according to the % of inertia they explain.

[17] For the PC's that represent a small part of inertia, nothing certifies a priori that the sixth PC in one PCA correspond to the sixth in another one.

extreme changes of regressors. For M4, all regressions include the first two factors and the sixth one, and at a large majority, the eighth one. On average, four or five PC's are present in the 72 regressions. All estimation errors are computed with the GDP flash estimate growth rates.

13.3.3 Evaluation and combination of forecasts (the GDP case)

Forecast evaluation

The preliminary step is to test the forecast unbiased-ness. This property is often rejected in practice especially with real time forecasts. We start by testing the sufficient condition of unbiased-ness. Table 13.4 gives the P-values of this test for the five competing models according to the estimation dates. The null hypothesis is largely accepted for all models except for the model M2, for which the P-values are equal to 3% (whatever the estimation dates). However, if we perform the simple test of unbiased-ness (Table 13.5), the forecasts of model M2 appear to be unbiased.

The second step consists in computing the root mean squared forecast errors (RMSFE) of all models and for all estimation dates. The results (in percentage point) are reported in Table 13.6 and provide a net classification of the five competing models. The best is M4, followed by M3, then M2 appears, followed by M1 and finally, the worst is M5. Consequently, the factor models (M3 and M4) dominate the bridge models (M1 and M2), and the two models which do not contain the IPI series (M2 and M5) are the less accurate. As all models give a large forecast error for the fourth quarter of 2006, we compute again the RMSFE without this quarter. The results are reported in Table 13.7. The five models are ranked similarly with lowest RMSFE: they vary between 0.11 percentage point and 0.22 (against 0.14 to 0.23 in Table 13.6).

Table 13.4: P-values of the test of the sufficient condition for unbiased-ness. Regression of the test $y_{t+1} = a_j + b_j \widehat{y}^j_{t+1,t} + \eta^j_{t+1}$ with null hypothesis $\{a_j = 0,\ b_j = 1\}$

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	8%	3%	33%	57%	78%
d_2: end of month 3 of quarter T	50%	3%	77%	89%	74%
d_3: end of month 1 of quarter $(T+1)$	12%	3%	43%	75%	71%

Table 13.5: P-values of the test of unbiased-ness. Regression of the test $y_{t+1} - \widehat{y}^j_{t+1,t} = a_j + \eta^j_{t+1}$ with null hypothesis $\{a_j = 0\}$

Estimation dates of the GDP of quarter T	M2
d_1: end of month 2 of quarter T	86%
d_2: end of month 3 of quarter T	79%
d_3: end of month 1 of quarter $(T+1)$	84%

Table 13.6: RMSFE (in percentage point) using real time data over 2002Q1-2007Q4 according to the estimation dates

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	0.20	0.22	0.17	0.16	0.23
d_2: end of month 3 of quarter T	0.17	0.22	0.15	0.14	0.23
d_3: end of month 1 of quarter $(T+1)$	0.18	0.22	0.16	0.14	0.23

Table 13.7: RMSFE (in percentage point) using real time data over 2002Q1-2007Q4 according to the estimation dates, without 2006Q4

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	0.19	0.22	0.17	0.15	0.21
d_2: end of month 3 of quarter T	0.15	0.22	0.1	0.12	0.21
d_3: end of month 1 of quarter $(T+1)$	0.15	0.21	0.14	0.11	0.22

The third step consists in verifying if M4 is the encompassing model. In this case, the other models are redundant and all the relevant information is contained in M4. We test – using regression (13.10) – the null

hypothesis $\{\alpha_4 = 1,\ \alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_5 = 0\}$. The P-values are equal to 94%, 74% and 97% depending on the estimation dates (Table 13.8). As the null hypothesis is largely accepted, this leads to conclude that M4 dominates all other models. Unfortunately, if now we perform the same test for M3 instead of M4, i.e. we test the null hypothesis $\{\alpha_3 = 1,\ \alpha_k = 0 \text{ for } k \neq 3\}$, we accept the dominance of M3 because the P-values are equal to 39%, 44% and 27% depending on the estimation dates (Table 13.8). Table 13.8 shows that, even for M1, the P-values of the null hypothesis $\{\alpha_1 = 1,\ \alpha_k = 0 \text{ for } k \neq 1\}$ exceed 5% (respectively equal to 8%, 15% and 7%). The null hypothesis of the encompassing test is only rejected for M2 and M5 (Table 13.8). Table 13.9 shows the P-values of the same tests when the fourth quarter of 2006 is omitted. The results are more in accordance with our expectations. The P-values are low except naturally for M4 and unfortunately for M3, but in this case they exceed 5% only for two estimation dates (where they are equal to 22% and 17%). To have more contrasted results, it would be necessary to implement the test on a larger sample (here, 24 forecasts per estimation dates are available). However, Table 13.9 suggests that M4 dominates the other models and, consequently, that combining will appear useless.

Table 13.8: P-values of the forecast encompassing test according to the estimation dates. Regression of the test $y_{t+1} = \alpha_0 + \alpha_1 \widehat{y}^1_{t+1,t} + \alpha_2 \widehat{y}^2_{t+1,t} + \ldots + \alpha_5 \widehat{y}^5_{t+1,t} + \epsilon_{t+1}$. Null Hypothesis for column Mj $\{\alpha_j = 1,\ \alpha_k = 0\ for\ k \neq j\}$

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	8%	2%	39%	94%	2%
d_2: end of month 3 of quarter T	15%	0%	44%	74%	0%
d_3: end of month 1 of quarter $(T+1)$	7%	1%	27%	97%	0%

Table 13.9: P-values of the forecast encompassing test according to the estimation dates without 2006Q4

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	4%	1%	22%	51%	1%
d_2: end of month 3 of quarter T	6%	0%	17%	70%	0%
d_3: end of month 1 of quarter $(T+1)$	0%	1%	3%	86%	0%

Forecast combination

We expect a priori that the gain of accuracy obtained by combining our forecasts will be low because four of the five models (the structural ones) provide highly positively correlated forecasts. This is inevitable because the available information to construct structural models is limited. Consequently, all models are based on a set of common series and produce positively correlated forecasts. Tables 13.10, 13.11 and 13.12 give the correlation coefficients between a pair of forecast errors for each date d_1, d_2 and d_3. In average, the correlation coefficient between a pair of structural forecast errors is equal to 0.85, whereas it falls below 0.50 between the time series forecasts and another structural forecast.

Nevertheless, we search for estimating optimal weights which sum to one over the 2002-2007 forecast period. These in-sample weights could be used in the future. If the weights are not constrained to be non negative, we obtain the weights [18] reported in Table 13.13. Some are negative, this is the consequence of the high level of correlation between forecasts. As we consider that weights have to be non negative, we look for optimal non negative weights which sum to one. These optimal weights[19] are listed in Table 13.14. We observe that the forecasts from M2 are excluded for the three estimation dates and those coming from M1, for two dates. Thus, practically only factor models have a priori to be combined as well as the time series model. The presence of the latter is explained by its low level of correlation with the other models.

To see if any combination gain could be expected in our context, we now use these weights to form the ideal combined forecasts, ideal because these weights are not reachable in practice since they are in-sample weights. This ideal combination is compared to three others:

[18] The hypothesis of weights summing to one and of no intercept (to correct bias) is tested and this gives P-values respectively equal to 88%, 79% and 99% for the three estimation dates.

[19] The P-values of the hypothesis specified in endnote 17 are equal respectively to 86%, 93% and 98%.

Table 13.10: Correlations between forecast errors (date d_1)

M1	M2	M3	M4	M5
1				
0.86	1			
0.93	0.86	1		
0.81	0.72	0.85	1	
0.49	0.27	0.53	0.53	1

Table 13.11: Correlations between forecast errors (date d_2)

M1	M2	M3	M4	M5
1				
0.73	1			
0.93	0.81	1		
0.83	0.61	0.78	1	
0.50	0.27	0.52	0.51	1

Table 13.12: Correlations between forecast errors (date d_3)

M1	M2	M3	M4	M5
1				
0.69	1			
0.89	0.72	1		
0.72	0.51	0.73	1	
0.37	0.25	0.47	0.51	1

Table 13.13: Optimal weights according to the estimation dates

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	-0.38	0.06	0.38	0.78	0.16
d_2: end of month 3 of quarter T	-0.52	-0.11	0.87	0.72	0.04
d_3: end of month 1 of quarter $(T+1)$	-0.16	0.07	0.27	0.75	0.07

Table 13.14: Optimal non negative weights according to the estimation dates

Estimation dates of the GDP of quarter T	M1	M2	M3	M4	M5
d_1: end of month 2 of quarter T	0	0	0.07	0.77	0.16
d_2: end of month 3 of quarter T	0	0	0.34	0.59	0.07
d_3: end of month 1 of quarter $(T+1)$	0.05	0	0.15	0.71	0.09

(i) equal weights

(ii) optimal weights when pair-wise correlations are set to zero (given by equation (13.7))

(iii) linearly decreasing weights according to the rank given by the RMSFE (5/15 for M4, 4/15 for M3, 3/15 for M1, 2/15 for M2 and 1/15 for M5).

The RMSFE of these four combinations are reported in Table 13.15 and compared to the RMSFE of M4. Only the ideal combined forecasts beat the forecasts of M4. But the gain would have been undetectable if the RMSFE (in percentage point) had been reported with two decimals instead of three! All others combinations imply a deterioration in comparison with the forecasts of M4. But they give comparable RMSFE showing that the weights have not much impact on the RMSFE.

To investigate another situation, we combine now the forecasts provided by M1, M2 and M5. The optimal unconstrained weights over the 2002-2007 forecast period are found to be positive. The optimal weights summing to one are reported in Table 13.16. They vary strongly with the estimation dates. These ideal combinations are compared to an equally-weighted combination. The RMSFE of these two combinations are listed in Table 13.17 and compared to the RMSFE of M1. We note that the two combinations give close RMSFE, showing the little impact of the weights. A combination gain (of order 0.02 percentage point) appears for two estimation dates. The weak reduction in forecast error is due to the high level of correlations between our forecasts. In this case no diversification gain can be expected.

Table 13.15: RMSFE (in percentage point) of four combined forecasts depending on the estimation dates. W1: equal weights. W2: $\alpha_j = \frac{\frac{1}{\sigma_j^2}}{\sum_{i=1}^{k}(\frac{1}{\sigma_i^2})}$. W3:linearly decreasing $\alpha_4 = \frac{5}{15}$ $\alpha_3 = \frac{4}{15}$ $\alpha_1 = \frac{3}{15}$ $\alpha_2 = \frac{2}{15}$ $\alpha_1 = \frac{1}{15}$. W4: optimal weights (see previous Table)

Estimation dates of the GDP of quarter T	W1	W2	W3	W4	RMSFE of M4
d_1: end of month 2 of quarter T	0.166	0.163	0.165	0.155	0.156
d_2: end of month 3 of quarter T	0.151	0.146	0.147	0.138	0.144
d_3: end of month 1 of quarter $(T+1)$	0.148	0.144	0.145	0.136	0.139

Table 13.16: Optimal non negative weights according to the estimation dates

Estimation dates of the GDP of quarter T	M1	M2	M3
d_1: end of month 2 of quarter T	0.19	0.36	0.45
d_2: end of month 3 of quarter T	0.75	0.01	0.23
d_3: end of month 1 of quarter $(T+1)$	0.49	0.18	0.33

Table 13.17: RMSFE (in percentage point) of two combined forecasts according to the estimation dates. W1: equal weights. W4: optimal weights (see previous Table)

Estimation dates of the GDP of quarter T	W1	W4	RMSFE of M1
d_1: end of month 2 of quarter T	0.177	0.175	0.198
d_2: end of month 3 of quarter T	0.166	0.160	0.167
d_3: end of month 1 of quarter $(T+1)$	0.162	0.160	0.180

13.3.4 Forecasting models of the monthly IPI growth rate

To forecast monthly IPI growth rate, we propose three models. Two models are structural regressions (P1, P2), the third one (P3) is the usual benchmark $AR(4)$ model. The (P1) model uses individual series as regressors, whereas the P2 model uses principal components. Each indicator $(j = 1, 2, 3)$ gives a one-month-ahead forecast $f^{(j)}_{T+1,T}$ of IPI for month $(T+1)$ just after the release of IPI for month T and, at the same date, a two-month-ahead forecast $f^{(j)}_{T+2,T}$ of IPI for month $(T+2)$.

The model (P1)

Table 13.18 presents the regressors entering the equation.

The factor model (P2)

The factors are extracted from a small data set whose composition is given by the LARS algorithm. The eight selected series are reported in Table 13.19. Beside the PC's, the regression model includes two autoregressive terms (lags 1 and 2) and a financial series, the real effective exchange rate growth rate (with 3 lags), exactly as it is in the model (P1).

The out-of-sample forecast errors done with this factor models using real-time data are computed over the 2002-2007 period. Thus, we perform 72 principal component analyses and 72 regressions. All regressions include the first two factors, and about a third, the third factor. All forecast errors are computed with the first estimate of the IPI growth rates.

With the same regressions, we compute two-month-ahead forecasts, adopting the following rule: when a survey data is missing, it is replaced by its previous value. It is certainly possible to do better, in principle extrapolations with $AR(p)$ models must be more accurate.

13.3.5 Evaluation and combination of forecasts (the IPI case)

Table 13.18: RMSFE (in percentage point) of two combined forecasts depending on the estimation dates. W1: equal weights. W4: optimal weights (see previous Table)

Series	Lag
Monthly industrial production growth rate	1
Monthly industrial production growth rate	2
Real effective exchange rate growth rate	3
Change in industrial confidence indicator	0
Change in industrial confidence indicator	1
Industrial confidence indicator	0

Table 13.19: Coincident and leading series used to construct PC of the IPI factor model according to their LARS rank at the end of 2007

Series	Lag
Change in opinion on assessment of export order books (industrial survey)	0
Change in opinion on unemployment expect. over the next 12 months (consumer survey)	0
Expected business situation (retail survey)	2
Change in industrial confidence indicator	1
Production expectations for the months ahead (industrial survey)	0
Change in opinion on production expectations for the months ahead (industrial survey)	1
Change in opinion on production trend in recent months (industrial survey)	0
Change in economic sentiment indicator	0

Forecast evaluation

First, the test of the forecast unbiased-ness is performed. The P-values are shown in Table 13.20. Two models (P1 and P2) deliver one-month-ahead biased forecasts. In both cases, forecasts over-estimate reality. Thus, the second step is to propose a real time correction of these biases. At the date T, the forecasts $f^{(j)}_{T+1,T}$ are produced $(j = 1, 2)$ and the previous forecasts $f^{(j)}_{T-s+1,T-s}$ for $s = 1 \dots (T-1)$ are available $(j = 1, 2)$. The real time correction of the biases is done according to:

$$\tilde{f}^{(j)}_{T+1,T} = f^{(j)}_{T+1,T} - \left(\frac{1}{T-1}\sum_{s=1}^{T-1} f^{(j)}_{T-s+1,T-s}\right) + \left(\frac{1}{T-1}\sum_{s=1}^{T-1} y_{T-s}\right)$$

At the date $(T+h)$, the correction is

$$\begin{aligned}\tilde{f}^{(j)}_{T+h+1,T+h} = f^{(j)}_{T+h+1,T=h} - \left(\tfrac{1}{T+h-1}\textstyle\sum_{s=1}^{T+h-1} f^{(j)}_{T+h-s+1,T+h-s}\right) \\ + \left(\tfrac{1}{T+h-1}\textstyle\sum_{s=1}^{T+h-1} y_{T+h-s}\right)\end{aligned} \tag{13.25}$$

In other words, one subtracts from the new forecast the mean of the previous available forecasts and one adds the mean of the series to forecast computed on the same period.

After doing these bias corrections for the forecasts of the models P1 and P2, tests of the sufficient condition of forecast unbiased-ness are performed. The P-values of these tests are shown in Table 13.21. The sufficient condition of unbiased-ness is accepted for the corrected forecasts of the models P1 and P2 and also, for the forecasts of all other models.

The RMSFE of all models are listed in Table 13.22. The ranking resulting from the RMSFE criterion, puts the factor model P2 in the first place, followed by model P1, the $AR(4)$ model P3 being the worst. For the two-month-ahead forecasts, the factor model beats the $AR(4)$, but not significantly. Thus, structural models are interesting only for the one-step-ahead horizon.

For this horizon, we want to know if the factor model P2 is the encompassing model. Table 13.23 suggests that it is. Thus, we presume that combining forecasts will be useless.

Table 13.20: P-values of the test of unbiased-ness
(Test given in Table 5)

Forecasting model (IPI)	P-value
P1 (one-month-ahead)	0.1%
P2 (one-month-ahead)	0.2%
P3 (one-month-ahead)	54%
P2 (two-month-ahead)	47%
P3 (two-month-ahead)	63%

Table 13.21: P-values of the test of the sufficient condition for unbiased-ness
(Test given in Table 4)

Forecasting model (IPI)	P-value
P1 (one-month-ahead), bias corrected	22%
P2 (one-month-ahead), bias corrected	33%
P3 (one-month-ahead)	57%
P2 (two-month-ahead)	45%
P3 (two-month-ahead)	60%

Table 13.22: RMSFE (in percentage point) using real time data over 2002M1-2007M12

Forecasting model (IPI)	RMSFE
P1 (one-month-ahead), bias corrected	0.57
P2 (one-month-ahead), bias corrected	0.55
P3 (one-month-ahead)	0.62
P2 (two-month-ahead)	0.64
P3 (two-month-ahead)	0.65

Table 13.23: P-values of the forecast encompassing test

Forecasting model (IPI)	P-value
P1 (one-month-ahead), bias corrected	8.8%
P2 (one-month-ahead), bias corrected	68.7%
P3 (one-month-ahead)	0.0%

Forecast combination

As previously, we expect that the gain of accuracy obtained by combining our forecasts will be low because the three models provide highly positively correlated forecast errors. Tables 13.24 and 13.25 give the correlation coefficients between all pair of forecast errors: 0.93 between the one-month-ahead forecasts of P1 and P2, less than 0.8 when one of the forecast is given by P3 (the $AR(4)$ model). The three models have in common two autoregressive terms, which explains the high correlations.

Nevertheless, as previously for the GDP, we search for estimating optimal weights which sum to one over the forecast period 2002-2007. These in-sample weights are the ideal ones and will indicate the maximum achievable combination gain. In spite of high level of correlations between forecasts, the unconstrained weights are positive. Imposing a sum equal to one is accepted and we obtain the weights reported in Table 13.26 (respectively 0.09 for P1, 0.71 for P2 and 0.20 for P3). The forecasts derived from P3 have a non neglected weight (0.2) due to their lower level of correlation with the other forecasts. However, these ideal weights do not lead to a real combination gain as shown in Table 13.27. The RMSFE of this combination is equal to 0.54 (against 0.55 with model P2). This ideal combination is compared with two others: (i) linearly decreasing weights which lead to a RMSFE equal to 0.54 (Table 27), (ii) equal weights which provide a RMSFE equal to 0.55. As previously, the weight role is negligible, all combinations giving no significantly different RMSFE.

For the two-month-ahead forecasts the ideal combination accords a weight equal to 0.58 for P2 and 0.42 for P3, close to the equally-weighted combination (Table 13.28). Their RMSFE are not significantly different from that of model P2 (Table 13.28).

Table 13.24: Correlations between forecast errors (one-month-ahead)

P1	P2	P3
1		
0.93	1	
0.81	0.80	1

Table 13.25: Correlations between forecast errors (two-month-ahead)

P1	P2
1	
0.93	1

Table 13.26: Optimal weights over period 2002-2007

Forecasting model (IPI)	weight
P1 (one-month-ahead), bias corrected	0.09
P2 (one-month-ahead), bias corrected	0.71
P3 (one-month-ahead)	0.20

Table 13.27: RMSFE (in percentage point) of three combined forecasts

Combined forecast (one-month-ahead)	RMSFE
Optimal in-sample weights	0.54
Linearly decreasing (1/6 for P3, 2/6 for P1, 3/6 for P2)	0.54
Equally-weighted	0.55

Table 13.28: RMSFE (in percentage point) of two combined forecasts

Combined forecast (two-month-ahead)	RMSFE
Optimal in-sample weights (0.58 for P2, 0.42 for P3)	0.634
Equally-weighted	0.635

13.3.6 Conclusion

The gain in combining rests on the well known diversification principle. When the forecasts to combine are highly positively correlated this principle does not work. This is what happens in our two examples. But it would probably occur for other PEEI series. The reason is simple: pertinent structural models are necessarily built on a common economic information, rather limited, generating an unavoidable similarity between forecasts. This is particularly true for the euro area where the number of available series is not large, especially if we only deal with real time data.

The property that a weighted mean is not much sensitive to its weights is perfectly illustrated in our examples. Rather different weight profiles have little consequences on the associated RMSFE. Linearly decreasing weights according to the ranking of the models with the RMSFE criterion seems lead to slightly better results than equal weights.

Combining forecasts is a mean of using more information to base forecasts. Instead of enlarging information by combination, it may be preferable to base forecasts directly on large information sets. It is the spirit of factor models. Here these models (M4 and P2) give the best results in both examples. They can be considered as dominant in the sense of the forecast encompassing test. Note that our factor models are not based on a very large data set, because following Bai and Ng (2008), we think that a smaller but well targeted data set is preferable. Introducing data not directly related to the target increases the forecast volatility by creating much noise for nothing.

Bibliography

Bai, J. and Ng, S. (2008), Forecasting Economic Time Series Using Targeted Predictors, Journal of Econometrics 146, pp. 304–317.

Bates, J., and Granger, C.W.J. (1969), The Combination of Forecasts, Operations Research Quarterly, 20, 451–468.

Boivin, J., and S. Ng, (2006), Are more data always better for factor analysis?, Journal of Econometrics, vol 132, issue 1, May, pp. 169–194.

Chan, Y.L., Stock, J.H., and Watson, M.W. (1999), A dynamic factor model framework for forecast combination, Spanish Economic Review, 1, pp. 91–121.

Charpin, F., Mathieu, C., and G.L. Mazzi (2008), Construction of coincident indicators for the euro area, 28th Annual International Symposium on forecasting.

Chong, Y.Y. and Hendry, D.F. (1986), Econometric evaluation of linear macroeconomic models, Review of Economic Studies, Vol. 53, pp.671–690.

Clemen, R.T. (1986), Linear Constraints and the Efficiency of Combined Forecasts, Journal of Forecasting, 5, pp. 31–38.

Clemen, R.T. (1989), Combining forecasts: A review and annotated bibliography, International Journal of Forecasting, 5, pp. 559–583.

Coulson, N. E., and Robbins, R. P. (1993), Forecast combination in a dynamic setting, Journal of Forecasting,12, pp. 63–67.

Diebold, F. X. (1988), Serial correlation and the combination of forecasts, Journal of Business and Economic Statistics, 6, pp. 105–111.

Diebold, F. X., and Lopez, J. A. (1996), Forecast evaluation and combination, in Maddala and Rao (eds), Handbook of Statistics, volume 14, pp. 241-268, North-Holland.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. 2004, Least Angle Regression, Annals of Statistics 32:2, pp. 407–499.

Ericsson, Neil R., (1992), Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extension, and Illustration, Journal of Policy Modeling 14 (4), pp. 465–495.

Ericsson, Neil R., and Marquez ,J., (1993), Encompassing the Forecasts of U.S. Trade Balance Models, The Review of Economics and Statistics, Vol. 75, pp. 19–31.

Figlewski, S., (1983), Optimal price forecasting using survey data, Review of Economics and Statistics, 65, pp. 13–21.

Granger, C.W.J., and, Newbold, P. (1977), Forecasting Economic Time series, Chapter 8, Academic Press.

Granger, C.W.J., and Ramanathan, R. (1984), Improved Methods of Forecasting, Journal of Forecasting, 3, pp. 197–204.

Hendry, D. F., and Clements, M. P., (2002), Pooling of Forecasts, Econometrics Journal, volume 5, pp. 1–26.

Nelson, C. R.(1972), The prediction performance of the FRB-MIT-PENN model of the US economy, American Economic Review, 62, pp. 902-917.

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

Timmermann, A. (2006), Forecast combinations, Handbook of Economic Forecasting, Vol. 1, Chapter 4, Elsevier.

14 Combined Nowcasts

Authors

Silvia Lui
University of Groningen
s.w.s.lui@rug.nl

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

14.1 Introduction

This chapter considers the potential for forecast combination to improve the accuracy of nowcasts or rapid estimates for the following five PEEIs measured using EU25 and EA12 data:

1. Industrial output price index for domestic markets (monthly);
2. Industrial new orders index (monthly);
3. Employment (quarterly);
4. Industrial turnover (monthly);
5. GDP (quarterly).

The empirical exercise, based on recursive out-of-sample simulations, involves comparison of the combined nowcasts against nowcasts/forecasts produced by univariate linear models with multiple indicators, where information criteria (such as the Bayesian Information Criterion, BIC) are used to select recursively the best-fitting model. The application to GDP, in contrast to the first four PEEIs, involves using real-time data from EUROIND, a real-time database, and is therefore not simply a *pseudo* real-time exercise and accommodates data revisions.

We find that combination can help improve nowcast accuracy, and that weighted combinations appear to work better than equal weighted combinations. We focus on nowcasting GDP growth 15 days after the end of the quarter (i.e., about 30 days ahead of Eurostat's official Flash estimate), using a parsimonous but widely used set of indicator variables available at the higher monthly frequency: industrial production, retail trade data and qualitative survey data. Publication lags for these monthly indicators mean that at 15 days, producing the quarterly rapid estimate for GDP growth involves forecasting the final month in the quarter for industrial production and retail trade - and then using the constructed quarterly aggregate to nowcast quarterly GDP growth. As a benchmark we also report results at 45 days, when the third month in the quarter (for industrial production) is known – of course this corresponds to the same horizon at which Eurostat currently produces its Flash estimate. Forecast accuracy is evaluated relative both to the first, second and third-release and the "final"-vintage GDP growth data. Thereby we analyse the ability of combination methods to circumvent structural instabilities and deal with data revisions.

14.2 Applications to Five PEEIs

As previous Chapters in this Handbook explain, rather than select a single 'best' forecast it can be an advantage to combine competing forecasts. This follows from appreciation of the fact that, although one forecasting model may be 'better' than the others, we may not select it with probability one; i.e., we may not be sure that it is the best forecast. Therefore, if we considered this single forecast alone we would be overstating its precision. We may better approximate the truth, and account for the uncertainty in model selection, by combining forecasts.

Indeed, it is well recognised both theoretically and empirically that combining competing individual point forecasts of the same variable can deliver more accurate forecasts, in the sense of a lower Root Mean Squared Error; see Bates and Granger (1969), Stock and Watson (2004) and Timmermann (2006). The success of combination follows from the fact that individual forecasts may be based on misspecified models, poor estimation or non-stationarities. Moreover, work such as Hendry and Clements (2004) explores further why point forecast combination works through analytical and Monte Carlo investigation.

The plan of the empirical work in this Chapter is as follows. In Section 14.2.1 we review the methods used to construct the competing point forecasts or nowcasts. This involves computing nowcasts from univariate linear models with multiple indicators. But rather than just use an information criteria (such as the BIC) to select recursively the best-fitting model, we compare the performance of this selected forecast with that

based on combining all possible nowcasting models. We consider both equal weighted combinations and BMA combinations (as reviewed in earlier Chapters in this Handbook) where the BIC is used as the penalty function and attaches a higher weight in the combination to forecasts from those models that better explain the data.

Then in section 14.2.1 we consider the use of monthly indicators to nowcast a quarterly variable, specifically GDP growth. This involves the use of bridge equations. But again the focus is to compare the selected forecast with the combined one.

Section 14.2.1 reviews the characteristics of the nowcasts produced for the five PEEIs. In particular, we note how quickly these nowcasts are available (ahead of Eurostat's current first estimates) and review the set of indicator variables available at this timetable.

Section 14.2.2 then compares the accuracy of selected and combined forecasts using recursive out of sample simulations using the latest data vintage for the industrial output price, industrial new orders, quarterly employment and industrial turnover. We evaluate the quality of Flash estimates for both (t/t-1) and (t/t-12) or (t/t-4), for quarterly data, transformations of the underlying series of interest.

Section 14.2.3 then considers the real-time application to quarterly GDP using real-time data as available in the EUROIND database. This is important, as it looks at the ability of combination to offset revisions to the official data. Clark and McCracken (2008), for example, in an application to real-time US data, found that combined forecasts worked well in the face of structural breaks and data subject to revisions.

14.2.1 Methodology for FLASH estimation: a review

The forecasting approach is designed to produce nowcasts (or rapid estimates) that are credible to policy-makers and other non-statisticians. We see this as ruling out processes with lengthy lags in exogenous variables, since it is difficult to defend a situation where an indicator is sharply influenced by some other variable up to six months or so ago. On the other hand it is perfectly possible to defend a structure in which changes in fuel prices affect output prices, for example, gradually through a partial adjustment process.

The desire to produce clear models with short lags is reinforced by the fact that in many cases the data series we have available are although monthly, generally short in duration. This means that it is difficult to explore co-integration satisfactorily and, at least in the first instance, our models will be regression equations constructed with the dependent variable entering only in log differences.

The modelling framework requires only a one period ahead forecast. This means that there is no distinction between single equation and multivariate models such as VAR models.[1]

Regression-based estimates are produced as special cases of the following general regression equation:

$$\Delta y_t = c + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{i=0}^{p} \sum_{j=1}^{k} \beta_{ij} x_{t-i,j} + u_t \; ; t = 1, 2, ..., T, \tag{14.1}$$

where y_t is the log of the dependent variable, $x_{t,j}$ is the j-th indicator variable ($j = 1, 2, ..., k$) in logs when appropriate, c is an intercept, p is the number of lags, Δ is the first-difference operator and u_t is a mean zero disturbance with variance σ^2. All indicator variables that enter (14.1), if necessary, are differenced until stationary. The use of (log) first differences, except for series (like the survey balances) which by definition are bounded and hence I(0), is deemed sufficient to render all series stationary. In the presence of such small sample periods this assumption is not tested (e.g. via a unit root test), since any test would have such low power as to be pretty meaningless. Nevertheless, there is a consensus that price and output variables are I(1), and we proceed on this (plausible) assumption and therefore first difference them.

[1] We also note that for short horizons the forecasting performance from univariate nonlinear models is typically worse or not much better; see Stock and Watson (1999) and Marcellino (2008). We therefore confine attention to simpler linear models. We also note there is a large literature using factor methods to produce nowcasts and forecasts.

Cointegrating restrictions could be imposed on (14.1) if cointegration is present. But there are a number of reasons for working with differenced (stationary) series rather than also including possible cointegrating relationships between integrated variables, aside from the small-sample we have in our application which would render any test for cointegration meaningless due to the low power; see Hoffman and Rasche (1996) and Diron (2008).

It should be noted that contemporaneous values of the indicator variables are included in (14.1). This reflects the fact that these indicators by their nature are published ahead of the variables to which they are assumed to relate, even though they may relate to the same time period.

Various specifications of (14.1) might be considered based on different methodological approaches to selecting the specification to use for regression-based estimation. We will work as follows:-

Given k indicator variables (our review of data suggests that k is unlikely to exceed 4) and a given number of lags (we consider $p = 2$), for $t = 1, ..., T$, we consider all possible combinations of (14.1) of the twelve exogenous and two lagged endogenous variables thus generated. Since, however, this creates a very large possible number of regressions and bearing in mind the well-known benefits of parsimony in forecasting models, we limit ourselves to those equations containing no more than four explanatory variables. There are $1 + 14 + {}^{14}C_2 + {}^{14}C_3 + {}^{14}C_4 = 1470$ such equations. We then "automatically" select the preferred model using the Bayesian Information Criterion (BIC). Use this model, and its estimated coefficients, and the time T values of the explanatory variables in the preferred model for time T. This provides the model we use to forecast for period $T+1$. We denote the total number of explanatory variables, whether lags of the dependent variable or contemporaneous and lagged values of the indicator variables, by s. We set $s = 2$ in this study. However, results are qualitatively similar for other values of s.

Monthly Bridge Equations when Nowcasting Quarterly Variables (like GDP growth)

Various methods are available to nowcast a quarterly variable like GDP exploiting monthly indicator variables. Here we consider the use of bridge equations, since this framework sits naturally with our focus on (small N) indicator-based nowcasts from (14.1) - the production of so-called rapid estimates using methodologies which are familiar to national accountants.

Bridging involves linking monthly data, typically released early in the quarter, with quarterly data like GDP; e.g. see Baffigi et al. (2004). In effect a two-equation system is used to nowcast Δy_t, with the second equation comprising the forecasting model for the monthly variable $x_{t,j}$.[2] In common with much previous work, see Diron (2008), we consider simple AR models for $x_{t,j}$:

$$\Delta y_t = c + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \sum_{i=0}^{p} \sum_{j=1}^{k} \beta_{ij} \widehat{x}_{t-i,j} + u_t \; ; t = 1, 2, ..., T, \tag{14.2}$$

$$x_{t,j} = B(L) x_{t-1,j} + e_{t,j} \tag{14.3}$$

The nowcasting model for Δy_t, (14.2), is therefore estimated (in-sample, $t = 1, ..., T$) using observed data on $x_{t,j}$. However, when wanting to nowcast Δy_{T+1} since we may only have partial information on $x_{T+1,j}$ (for some indicator variables, j) the predicted values $\widehat{x}_{T+1,j}$ from the AR model are used instead in (14.2). For example, when wanting to nowcast GDP growth earlier than 45 days after the end of the quarter, say at 30 days, there is only two months of data on industrial production available. The final month in the quarter is therefore forecast using the AR model. This forecasted value is then combined with the two months of observed data to obtain $\widehat{x}_{T+1,j}$.

To ensure that $\widehat{x}_{T+1,j}$ is transformed in a manner consistent with the quarterly variable Δy_t (which represents quarterly growth at a quarterly rate) having forecast the final month in the quarter we transform $\widehat{x}_{T+1,j}$ so that it represents quarterly growth, rather than the average of the monthly growth rate.

[2] The errors between the two equations are assumed orthogonal so that the equations are estimated separately.

This is achieved, for example following Mariano and Murasawa (2003). Consider the monthly variable in levels (rather than log differences) $z_{t,m}$, where the subscript t indicates the particular quarter and m the month within that quarter, $m = 1, 2, 3$, $t = 1, ..., T$. Then,

$$x_t = \log z_t - \log z_{t-1} = \frac{1}{3}\Delta \log z_{t,3} + \frac{2}{3}\Delta \log z_{t,2} + \Delta \log z_{t,1} + \frac{2}{3}\Delta \log z_{t-1,3} + \frac{1}{3}\Delta \log z_{t-1,2} \quad (14.4)$$

where $\Delta \log z_{t,3}$ is monthly growth.

As discussed above combination offers a means of integrating out model uncertainty, in other words of insuring ourselves against having picked the wrong (regression) model. There is a considerable body of work that has found forecast combination to often work well. Indeed equal weighting is often found to work as well as more complex (optimal - variance weighted) alternatives.

Therefore, we consider the benefits of combining the nowcasts from, say, the 1470 regressions. That is, rather than use the BIC to select the best-fitting model we combine all 1470 nowcasts.

There is a trade-off between the timeliness and accuracy of nowcasts

Nowcasts or rapid estimates can always be produced more quickly, but the question is whether the loss in accuracy which this typically involves is acceptable to the data producer.

- We consider nowcasting the industrial output price 12 days ahead of Eurostat's first estimate. The set of indicator variables considered, allowing for up to 2 lags of each, including their contemporaneous values, where $s = 2$, are qualitative survey data, HICP, energy prices, unit values and German industrial price inflation.
- We consider nowcasting the industrial new orders index 18-20 days ahead of Eurostat's first estimate. The set of indicator variables considered, allowing for up to 2 lags of each, including their contemporaneous values, where $s = 2$, are qualitative survey data and German industrial new orders.
- We consider nowcasting quarterly employment 30 days ahead of Eurostat's first estimate. The set of indicator variables considered, allowing for up to 2 lags of each, including their contemporaneous values, where $s = 2$, are qualitative survey data, German and French employment data and monthly unemployment data (which in fact when the nowcast is made is available for all 3 months in a calendar quarter).
- We consider nowcasting industrial turnover 12 days ahead of Eurostat's first estimate. The set of indicator variables considered, allowing for up to 2 lags of each, including their contemporaneous values, where $s = 2$, are qualitative survey data, German industrial turnover and the industrial output price.
- We consider nowcasting quarterly GDP 30 days ahead of Eurostat's first (Flash) estimate, i.e. at 15 days after the end of the quarter of interest. The set of indicator variables considered, allowing for up to 2 lags of each, including their contemporaneous values, where $s = 2$, are qualitative survey data, monthly industrial production data, with the third month in the quarter forecast using an AR, and monthly retail trade data where again the third month in the quarter is forecast.

We should note, that we did also consider combination forecasts with a far larger set of indicator variables, not the subset considered above which have been selected carefully using our priors about what might constitute useful indicators. Consideration of this larger set of indicators led to worse performing combination forecasts; the main characteristic being consideration of this larger set of indicator led to a far less volatile combined forecast. This is to be expected, as many of these additional indicators offer little or no explanatory power for the PEEI of interest. Accordingly including them in the combination essentially means we are forcing the combination forecast to the (in-sample) mean of the PEEI. Combination using a subset of forecasts, or trimming, as discussed above, is in fact a well known means of improving the performance of combination forecasts. We therefore focus on combination using this subset of indicators. But allowing for up to 2 lags of each indicator, means we nevertheless combine a large number of nowcasts.

We now consider two sets of applications designed to evaluate the likely benefits of combination when nowcasting PEEIs, the first to the first four indicators listed above (Industrial output price index, Industrial new orders index, Employment and Industrial turnover) using final vintage data (a pseudo real-time exercise), and the second using real-time GDP data.

14.2.2 Recursive simulations for the first four PEEIs

Tables 14.1-14.4 summarise the performance of BIC recursively (so as if in real-time) selected and combined (both equal weighted, EW, and BMA BIC weighted) nowcasts by reporting their bias against the outturn, the standard deviation of the point forecasting error, the MAE and the RMSE. We also report $Prop$, which is the proportion of N models (N denotes how many individual indicator-based nowcasts are produced) beaten by the combined forecast.

Table 14.1 shows that for the industrial output price, whether t/t-1 or t/t-12 transformed, combination does not help. The RMSE is lower for the forecast produced by the model which is selected as the best-fitting model by the BIC. However, the weighted combination does beat the equal weighted combination; this is seen in a lower RMSE and a higher proportion of individual models being beaten. Figures 14.1 and 14.2 provide a graphical indication of the performance of the selected and combined nowcasts for the industrial output price for both the EU25 and EA12. It shows that the BIC selected nowcast is more volatile and also tracks the outturn better. In contrast, the combined nowcasts is too smooth. While it picks up the general tendency in which the outturn moves, it is slow to react and therefore one can experience large point forecast errors from time to time.

A similar story holds for industrial turnover in Table 14.2 and Figures 14.3-14.4 (for the EU25 and EA12 respectively). Similarly in Table 14.3 and Figures 14.5-14.6 for industrial new orders, again selection beats combination. While the combined forecast typically beats over 75% of the individual forecasts, it does not beat the best individual forecast as selected, recursively, by the BIC.

However, for quarterly employment a different picture emerges; see Table 14.4. The RMSE of the equal weighed combination (EW) is similar or perhaps slightly better than the BIC selected nowcast. But the BMA combination delivers a lower RMSE than the selected nowcast. This is reflected by the fact that the BMA combination beats a high proportion of individual model nowcasts - indeed it beats all of them for the EA12. This improvement for the combination nowcast for employment is also seen in Figures 14.7-14.8. While the combination nowcast is still smoother than the selected nowcast, it does not make such costly mistakes when a mistake is made. For example, in Figure 14.8 the selected nowcast appears to make a big forecasting error in early 2007, gambling that growth was going to rise. This gamble was not correct and by producing a more moderate nowcast the combination is rewarded. In this sense combination does provide a useful hedge against model uncertainty. But it does appear that aside from the odd costly mistake forecast selection does deliver a nowcast which better tracks employment growth - again Figures 14.7-14.8 show that the combination nowcasts are much smoother than both the outturn and the selected nowcast.

Table 14.1: Industrial output price: performance of selected and combined forecasts from 2003m9-2008m8

EU25	Ind Output Price					
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.113	0.256	0.228	0.278		276
EW combo	0.182	0.455	0.386	0.487	0.84	276
BMA combo	0.169	0.416	0.354	0.446	0.84	276
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.306	0.570	0.516	0.643		276
EW combo	1.032	1.290	1.425	1.643	0.641304	276
BMA combo	0.658	0.822	0.870	1.047	0.89855	276
EA12						
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.060	0.129	0.113	0.141		276
EW combo	0.128	0.353	0.302	0.373	0.84	276
BMA combo	0.109	0.287	0.249	0.305	0.84	276
EA12						
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.240	0.370	0.370	0.438		276
EW combo	0.549	0.813	0.813	0.974	0.706522	276
BMA combo	0.373	0.494	0.491	0.616	0.9021	276

Table 14.2: Industrial turnover: performance of selected and combined forecasts from 2004m1-2008m8

EU25	Ind Turnover					
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.075	1.392	1.098	1.382		120
EW combo	0.121	1.851	1.473	1.839	0.8	120
BMA combo	0.115	1.813	1.442	1.801	0.86	120
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.294	1.916	1.567	1.922		120
EW combo	0.158	2.445	1.832	2.429	0.875	120
BMA combo	0.089	2.233	1.699	2.215	0.875	120
EA12						
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.116	1.585	1.275	1.575		120
EW combo	0.181	2.063	1.659	2.052	0.758	120
BMA combo	0.171	2.023	1.622	2.012	0.758	120
EA12						
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.302	1.839	1.436	1.847		120
EW combo	0.166	2.440	1.881	2.423	0.875	120
BMA combo	0.002	2.164	1.677	2.145	0.875	120

Table 14.3: Industrial new orders index: performance of selected and combined forecasts from 2004m1-2008m8

EU25	Ind New Orders					
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.079	1.393	1.098	1.382		210
EW combo	0.141	1.881	1.494	1.869	0.814	210
BMA combo	0.137	1.848	1.467	1.837	0.829	210
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.313	1.892	1.545	1.901		210
EW combo	0.458	2.449	1.869	2.470	0.905	210
BMA combo	0.244	2.278	1.733	2.271	0.905	210
EA12						
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.121	1.584	1.275	1.575		210
EW combo	0.212	2.099	1.702	2.091	0.814	210
BMA combo	0.203	2.065	1.669	2.056	0.814	210
EA12						
t/t-12	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.318	1.799	1.413	1.811		210
EW combo	0.422	2.583	2.040	2.595	0.843	210
BMA combo	0.163	2.329	1.827	2.314	0.905	210

Table 14.4: Quarterly employment: performance of selected and combined forecasts from 2002q3-2008q2

EU25	Employment					
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.006	0.107	0.083	0.105		190
EW combo	0.042	0.101	0.090	0.108	0.866	190
BMA combo	0.038	0.093	0.080	0.099	0.866	190
t/t-4	bias	sd	MAE	RMSE	Prop	N
BIC selected	0.013	0.232	0.192	0.227		190
EW combo	0.134	0.156	0.161	0.203	0.914	190
BMA combo	0.102	0.146	0.145	0.175	0.981	190
EA12						
t/t-1	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.011	0.159	0.113	0.156		190
EW combo	-0.002	0.133	0.108	0.131	0.990	190
BMA combo	0.000	0.132	0.104	0.129	1.000	190
EA12						
t/t-4	bias	sd	MAE	RMSE	Prop	N
BIC selected	-0.015	0.200	0.152	0.196		190
EW combo	-0.058	0.197	0.158	0.202	0.943	190
BMA combo	-0.049	0.176	0.149	0.179	1.000	190

Figure 14.1: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EU25 industrial output price inflation (both t/t-1 and t/t-12)

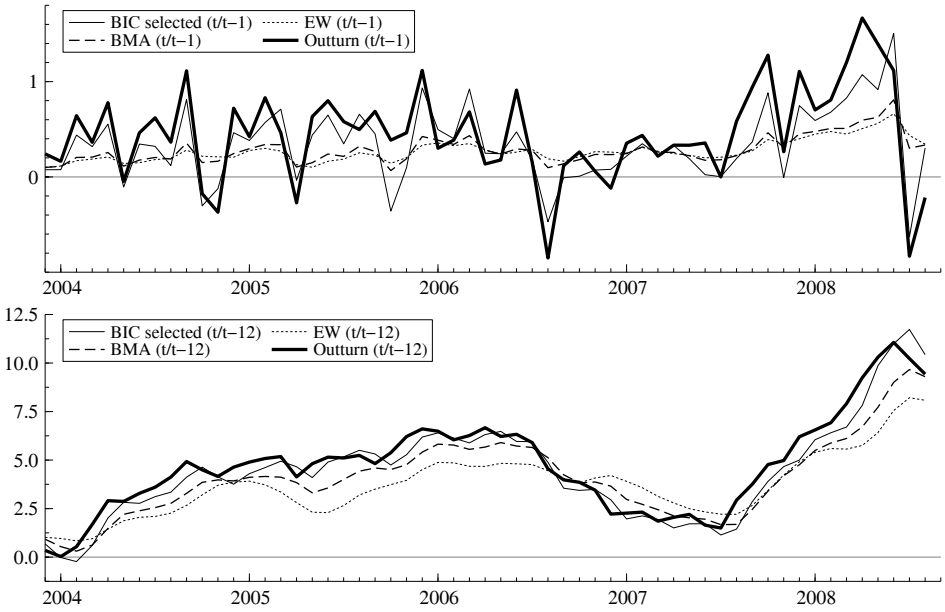


Figure 14.2: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EA12 industrial output price inflation (both t/t-1 and t/t-12)

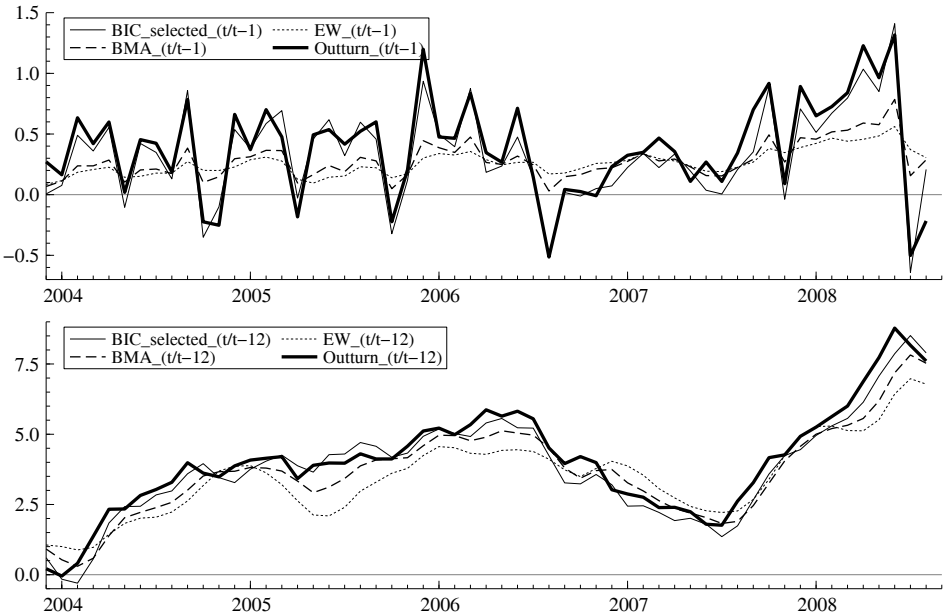


Figure 14.3: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EU25 industrial turnover index (both t/t-1 and t/t-12)

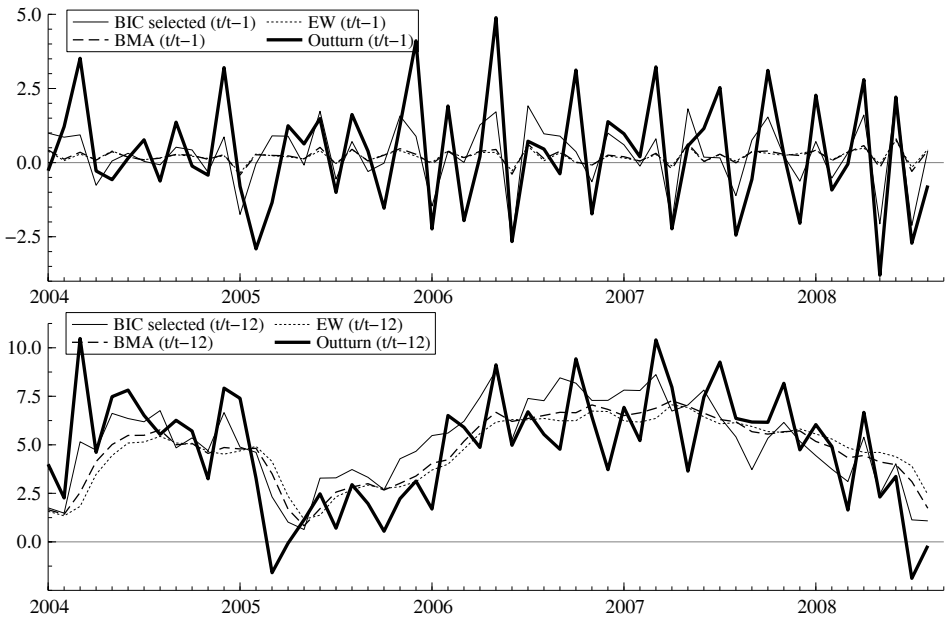


Figure 14.4: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EA12 industrial turnover index (both t/t-1 and t/t-12)

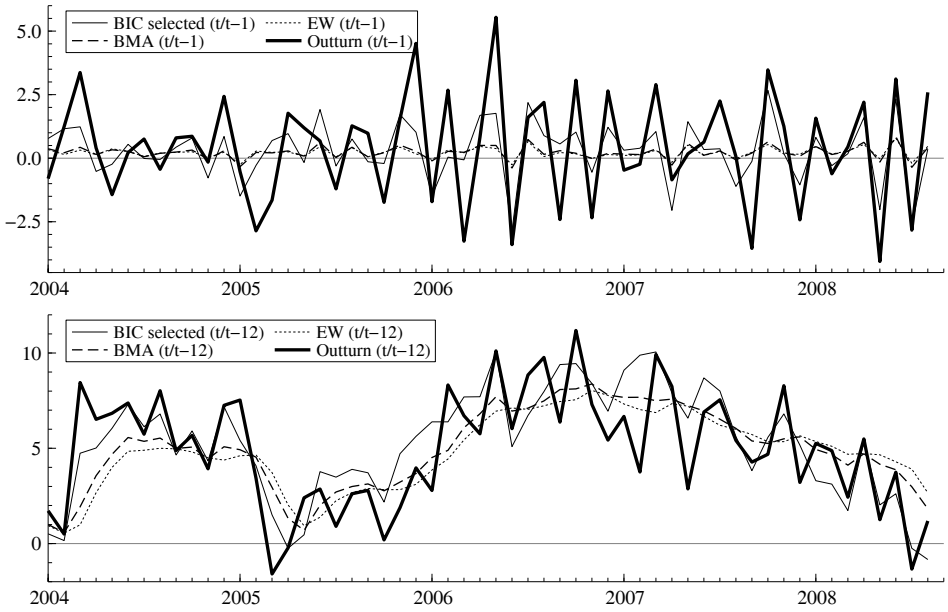


Figure 14.5: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EU25 industrial new orders index (both t/t-1 and t/t-12)

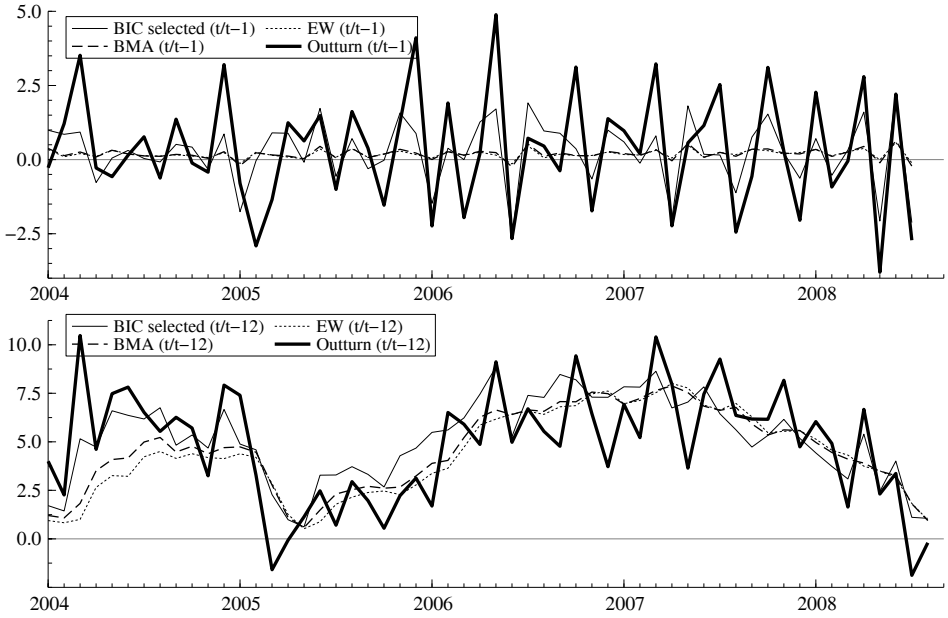


Figure 14.6: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EA12 industrial new orders index (both t/t-1 and t/t-12)

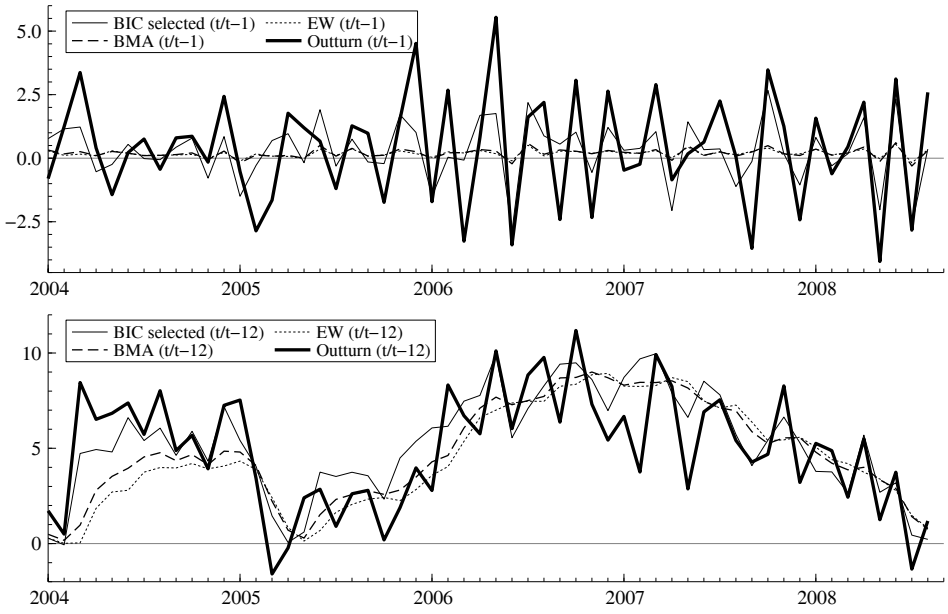


Figure 14.7: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EU25 employment growth (both t/t-1 and t/t-4)

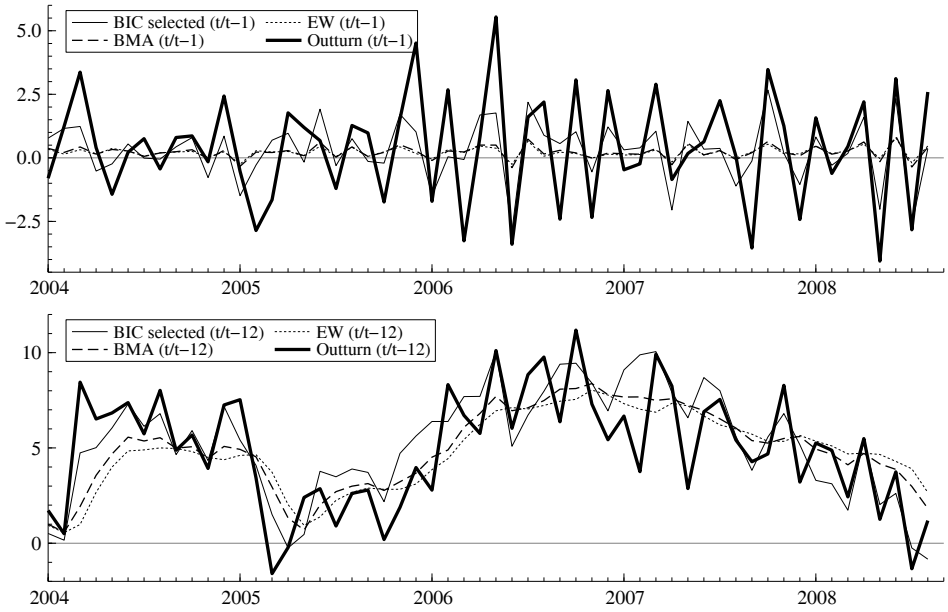
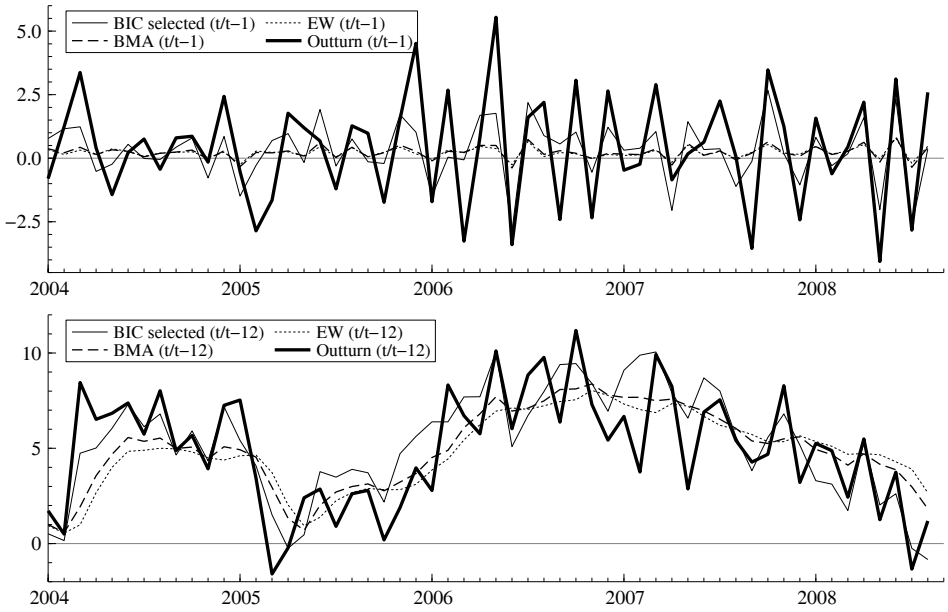


Figure 14.8: BIC selected, Equal Weighted combination and BMA combination Flash estimates of EA12 employment growth (both t/t-1 and t/t-4)



14.2.3 Real-time application to GDP growth using EUROIND database

In this section we evaluate the quality of combined and selected nowcasts for GDP growth at 15 days using the EUROIND database, and also make a comparison at 45 days, after the end of the quarter. Diron (2008) has also assessed the reliability of short-term forecasts of Eurozone GDP growth using (four years of) vintage data but does not consider model combination. Our analysis also exploits a longer sample.

Accordingly, from the EUROIND database we extract successive monthly vintages of EA12 GDP and EA12 industrial production (IP), a key indicator for GDP available at 45 days. We focus on GDP in constant prices.[3]

IP is typically available around the 10th-20th of each month, and vintages are available from May 2001. Until May 15 2003, the first GDP estimates for a given quarter were available at around 60 days after the end of the quarter. But from May 15 2003 Eurostat produced its first, so-called Flash estimate, at 45 days. We did also extract the real-time data for the (deflated) retail trade index; but a given vintage of data goes back to 1996 only.[4] Use of these retail trade data would, therefore, shorten our sample period too much - the loss of degrees of freedom would prevent a meaningful out-of-sample analysis. Moreover, using final vintage data we found that the retail trade index was of little or no use when nowcasting.

As an additional indicator to IP we also consider DG-ECFIN qualitative survey data. Specifically we consider the following balance statistics: (a) Industrial confidence indicator (40%); (b) Services confidence indicator (30 %); (c) Consumer confidence indicator (20%); (d) Retail trade confidence indicator (5%); (e) Construction confidence indicator (5%) and (f) The Economic sentiment indicator which is a composite measure.

To reflect the introduction of Eurostat's Flash estimates (at 45 days) for GDP in May 2003 we confine our out-of-sample analysis to the period post May 2003. Therefore, we use monthly vintages from May 2003 to December 2007. We focus on nowcasting GDP at 15 days. This involves, as above, for each successive monthly vintage of monthly IP data, forecasting the final month in the quarter using an AR model. The monthly estimates are then aggregated to a quarterly frequency as in (14.4), and a quarterly rate of growth, which is then related to quarterly GDP growth via the regression (14.1).

Tables 14.5-14.6 compare the performance of Flash (BIC selected) and combination nowcasts at 15 days, after the end of the quarter, for $s = 1$ and $s = 2$. Tables 14.7-14.8 consider their performance at 45 days, specifically when the final month in the quarter of IP does not have to be forecast. As a benchmark, we also report the performance of nowcasts which exploit information only on the qualitative survey data from DG-ECFIN. When $s = 2$ we combine 92 models. When $s = 1$ there are 13 models. In both cases these include AR models, with up to 3 lags.

First, comparison of Tables 14.5-14.6 and Tables 14.7-14.8, indicates that using the real-time data, there is little difference in accuracy as measured by the RMSE between the nowcasts computed at 15 or 45 days, irrespective of whether the nowcasts are combined or selected. This is perhaps a little surprising, and does contrast results using final vintage GDP data (not reported), where the RMSE does drop when we have three months of hard IP data available. Clearly the sample size is small - but revisions to the official data could also be playing a role.

Secondly, comparing the accuracy of selected versus combined nowcasts indicates, as with employment, that combination appears to help, with weighted combinations (BMA) having the slight edge over equal weighted combinations (EW). The combination nowcasts for GDP growth also beat the nowcasts formed only using qualitative survey data.

Thirdly, the nowcasts are more accurate when evaluated relative to the first GDP release from Eurostat rather than the latest (*final*) release.

[3] For GDP, there are vintages up to September 2008 in Current Prices (CP) but only November 2007 in Constant Prices (KP95). From November 2007, Eurostat did not produce data in KP95 anymore, but only chain-linked data (CLV2000). However, the chain-linked data do not go back into the past sufficiently for a meaningful real-time analysis so we consider KP95.

[4] There were some temporary changes in the publication calendar for retail trade in mid 2003, with Eurostat apparently experimenting with bringing forward the publication from 65 to 35 days. But only from the 5/3/04 vintage are these changes permanent.

Table 14.5: Comparison of BIC selected GDP growth nowcasts versus combined nowcasts for m=1 at 15 days: 2003q3-2007q4

		bias	sd	MAE	RMSE
BIC selected	First	-0.031	0.226	0.194	0.221
	Second	-0.033	0.232	0.198	0.228
	Third	-0.019	0.235	0.207	0.229
	Final	0.033	0.230	0.195	0.226
EW combo	First	-0.034	0.207	0.176	0.204
	Second	-0.036	0.207	0.174	0.204
	Third	-0.022	0.216	0.176	0.210
	Final	0.030	0.218	0.181	0.213
BMA combo	First	-0.033	0.205	0.177	0.201
	Second	-0.035	0.205	0.175	0.202
	Third	-0.021	0.213	0.176	0.208
	Final	0.031	0.215	0.179	0.211
Survey	First	-0.021	0.220	0.187	0.214
	Second	-0.024	0.217	0.186	0.212
	Third	-0.009	0.225	0.185	0.218
	Final	0.042	0.236	0.193	0.233

Table 14.6: Comparison of BIC selected GDP growth nowcasts versus combined nowcasts for m=2 at 15 days

		bias	sd	MAE	RMSE
BIC selected	First	-0.027	0.227	0.201	0.222
	Second	-0.029	0.230	0.199	0.225
	Third	-0.015	0.233	0.198	0.226
	Final	0.037	0.235	0.195	0.231
EW combo	First	-0.019	0.205	0.178	0.200
	Second	-0.022	0.204	0.175	0.199
	Third	-0.007	0.212	0.175	0.206
	Final	0.044	0.217	0.182	0.215
BMA combo	First	-0.022	0.202	0.179	0.197
	Second	-0.024	0.201	0.176	0.197
	Third	-0.010	0.209	0.176	0.203
	Final	0.042	0.213	0.178	0.211
Survey	First	-0.021	0.220	0.187	0.214
	Second	-0.024	0.217	0.186	0.212
	Third	-0.009	0.225	0.185	0.218
	Final	0.042	0.236	0.193	0.233

Table 14.9 lists the nowcasts (quarterly GDP growth in percentage points) from the BIC selected and two combination (EW and BMA) methods, computed in real-time using successive vintages from the EUROIND real-time data triangle, alongside the first, second, third and final official release of the data from Eurostat. Figure 14.9 plots these nowcasts, for $s = 2$, for the selected and combination forecasts alongside the first official estimate of GDP growth. We can see that all of the nowcasts track GDP quite well; and there is no obvious pattern suggesting the combination forecasts are either better or worse than the nowcast (based on selection).

But inspection of Figure 14.9 (top panel) does reveal that, as with the other four PEEIs, combination delivers smoother nowcasts than model selection and this can be beneficial since the selected nowcast makes the odd costly mistake. In contrast the combination nowcasts, which are quite similar, do track the behaviour of the official data well, although they do not pick up higher frequency movements.

Figure 14.9 also shows that there appears to be no obvious pattern to the revision between successive estimates of GDP growth. One cannot reject the null hypothesis that the revision series is mean zero.

Table 14.7: Comparison of BIC selected GDP growth nowcasts versus combined nowcasts for m=1 at 45 days

		bias	sd	MAE	RMSE
BIC selected	First	-0.039	0.234	0.202	0.230
	Second	-0.041	0.240	0.204	0.237
	Third	-0.027	0.243	0.214	0.238
	Final	0.025	0.239	0.198	0.234
EW combo	First	-0.034	0.207	0.177	0.203
	Second	-0.036	0.206	0.175	0.203
	Third	-0.022	0.215	0.177	0.210
	Final	0.030	0.217	0.180	0.213
BMA combo	First	-0.033	0.205	0.179	0.201
	Second	-0.035	0.204	0.177	0.201
	Third	-0.021	0.213	0.178	0.207
	Final	0.031	0.214	0.178	0.210

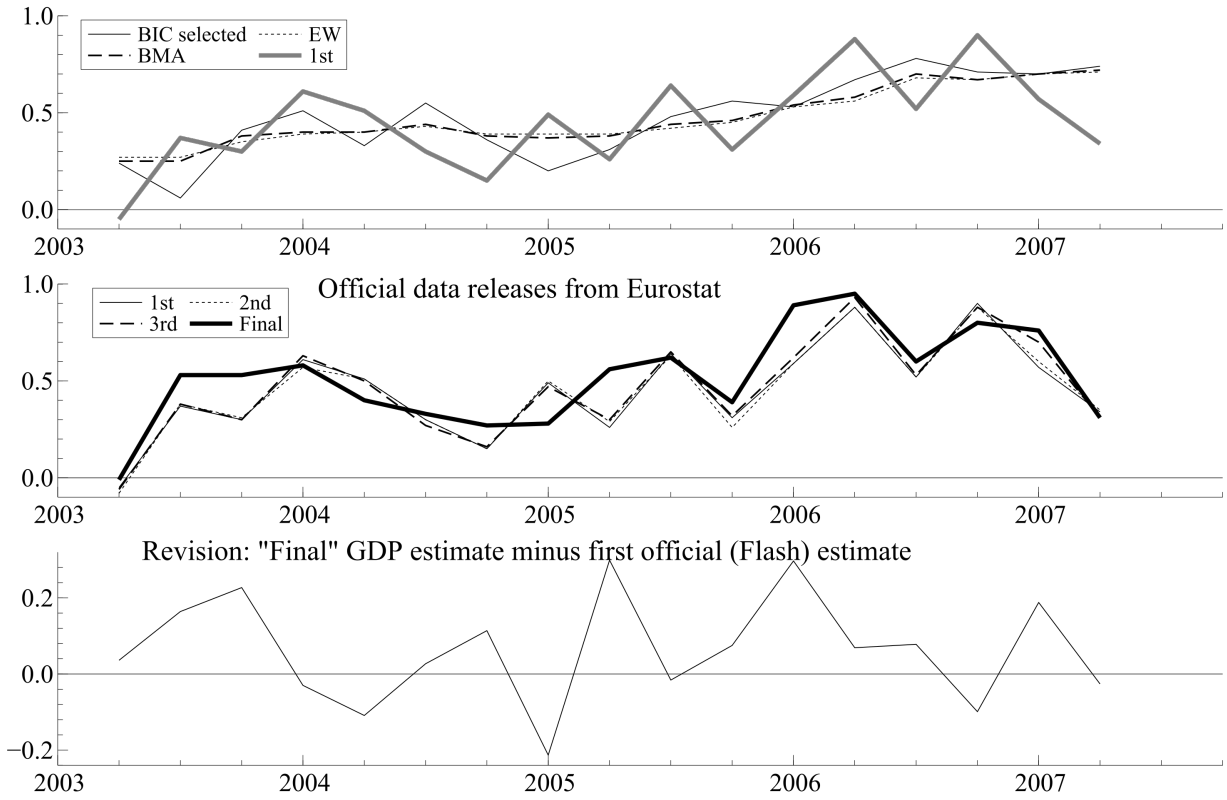
Table 14.8: Comparison of BIC selected GDP growth nowcasts versus combined nowcasts for m=2 at 45 days

		bias	sd	MAE	RMSE
BIC selected	First	-0.029	0.239	0.211	0.233
	Second	-0.032	0.242	0.209	0.237
	Third	-0.017	0.245	0.209	0.238
	Final	0.035	0.245	0.203	0.240
EW combo	First	-0.019	0.204	0.179	0.199
	Second	-0.022	0.204	0.177	0.199
	Third	-0.007	0.212	0.177	0.206
	Final	0.044	0.216	0.181	0.215
BMA combo	First	-0.021	0.203	0.182	0.198
	Second	-0.024	0.202	0.179	0.198
	Third	-0.009	0.210	0.179	0.204
	Final	0.042	0.214	0.178	0.211

Table 14.9: Flash, Equal Weighted combination forecasts and BMA combination forecasts for GDP growth, alongside the first, second, third and "final" release of quarterly GDP growth

	m=1			m=2			Outturn			
	Flash	EW	BMA	Flash	EW	BMA	1st	2nd	3rd	Final
2003q2	0.31	0.34	0.32	0.24	0.27	0.25	-0.05	-0.08	-0.06	-0.01
2003q3	0.12	0.33	0.31	0.06	0.27	0.25	0.37	0.38	0.38	0.53
2003q4	0.49	0.37	0.38	0.41	0.35	0.38	0.30	0.31	0.30	0.53
2004q1	0.59	0.40	0.41	0.51	0.39	0.40	0.61	0.57	0.63	0.58
2004q2	0.36	0.42	0.42	0.33	0.40	0.40	0.51	0.51	0.50	0.40
2004q3	0.59	0.44	0.45	0.55	0.43	0.44	0.30	0.30	0.27	0.33
2004q4	0.38	0.42	0.41	0.36	0.39	0.38	0.15	0.15	0.16	0.27
2005q1	0.19	0.42	0.40	0.20	0.39	0.37	0.49	0.50	0.47	0.28
2005q2	0.32	0.44	0.42	0.31	0.39	0.38	0.26	0.29	0.30	0.56
2005q3	0.53	0.45	0.46	0.48	0.42	0.44	0.64	0.63	0.65	0.62
2005q4	0.62	0.48	0.48	0.56	0.45	0.46	0.31	0.26	0.32	0.39
2006q1	0.53	0.53	0.53	0.53	0.53	0.54	0.59	0.59	0.62	0.89
2006q2	0.69	0.56	0.57	0.67	0.56	0.58	0.88	0.88	0.93	0.95
2006q3	0.73	0.64	0.65	0.78	0.68	0.70	0.52	0.52	0.53	0.60
2006q4	0.61	0.66	0.66	0.71	0.67	0.67	0.90	0.88	0.88	0.80
2007q1	0.56	0.68	0.68	0.70	0.70	0.70	0.57	0.60	0.70	0.76
2007q2	0.62	0.70	0.70	0.74	0.71	0.72	0.34	0.35	0.31	0.31
S.D.	0.17	0.12	0.13	0.21	0.15	0.15	0.24	0.24	0.25	0.25

Figure 14.9: Real-time Flash estimates for quarterly GDP growth using the EUROIND database



14.3 Conclusion

We consider the empirical performance of combination methods to produce nowcasts for five PEEIs. We contrast the combined nowcasts with those nowcasts based on selecting, recursively using the BIC, the preferred model from the set of N models considered.

We find that the ability of combination methods to produce more accurate nowcasts varies across PEEIs. For the monthly series, i.e. the industrial output price, industrial new orders and industrial turnover, the BIC selected nowcast is more volatile and better tracks the outturn. In contrast, the combined nowcasts is too smooth - as might be expected if some of the indicators used to nowcast have a weak relationship with the PEEI of interest.

But when nowcasting the quarterly PEEIs, namely employment and GDP growth, using higher-frequency monthly indicators, a different picture emerges. The combined nowcasts are now more accurate than the BIC selected nowcast. This is consistent with the indicators now having a stronger relationship with the PEEI of interest. There is also evidence that weighted combinations can outpredict equal-weighted combinations when nowcasting, constrasting the conventional wisdom (cf. Smith and Wallis (2009)) when combining forecasts.

Bibliography

Baffigi, A., Golinelli, R. and Parigi, G. (2004), Bridge models to forecast the euro area GDP, International Journal of Forecasting (20), 447-460.

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

Clark, T. E. and McCracken, M. W. (2008), Averaging forecasts from VARs with uncertain instabilities, Journal of Applied Econometrics . Forthcoming. Revision of Federal Reserve Bank of Kansas City Working Paper 06-12.

Diron, M. (2008), Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data, Journal of Forecasting 27(5), 371-390.

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

Hoffman, D. L. and Rasche, R. H. (1996), Assessing forecast performance in a cointegrated system, Journal of Applied Econometrics 11(5), 495-517.

Marcellino, M. (2008), A linear benchmark for forecasting GDP growth and inflation?, Journal of Forecasting 27(4), 305-340.

Mariano, R. S. and Murasawa, Y. (2003), A new coincident index of business cycles based on monthly and quarterly series, Journal of Applied Econometrics 18(4), 427-443.

Smith, J. and Wallis, K. F. (2009), A Simple Explanation of the Forecast Combination Puzzle, Oxford Bulletin of Economics and Statistics, Vol. 71, No 3, pp. 331-355.

Stock, J. H. and Watson, M. W. (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R. Engle and R. White, eds, 'Cointegration, causality and forecasting: a festschrift in honor of clive W.J. Granger', Oxford: OUP, pp. 1-44.

Stock, J. and Watson, M. (2004), Combination forecasts of output growth in a sevencountry data set, Journal of Forecasting 23, 405-430.

Timmermann, A. (2006), Forecast combinations, in G. Elliott, C. W. J. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting Volume 1', North-Holland, pp. 135-196.

VI Aggregate versus Disaggregate Approaches

15 Aggregate versus Disaggregate Approaches to Constructing Rapid Estimates

Authors

Silvia Lui
University of Groningen
s.w.s.lui@rug.nl

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

15.1 Introduction

Rapid estimates and/or forecasts can be computed using many different modelling approaches (as discussed in this handbook), in particular when interest rests with an aggregate variable, such as Euro-area GDP growth or a supranational aggregate (e.g. for the G20, OECD, etc.). Indeed most economic data involve aggregation over some dimension, such as over regions (within a country) and/or sectors. This chapter considers the pros and cons of deriving rapid estimates using the underlying disaggregated information, whether that be countries, regions, sectors or some other dimension, versus using aggregated data only.

The so-called aggregate or direct approach involves computing rapid estimates at the aggregate level, with an aggregate model adopted and an aggregate estimate produced directly by conditioning on aggregate indicators and/or lags of the aggregate itself. In contrast, the so-called disaggregate or indirect approach involves producing rapid estimates or forecasts at the disaggregate level, say the sectoral, regional or country level, using a disaggregate model which conditions on disaggregate information and then aggregates up to produce a nowcast or forecast for the aggregate. Alternatively, one might employ a mixed approach, in which an aggregate estimate is produced by conditioning on both the aggregate and disaggregate indicators (cf. Hendry and Hubrich (2011)). The existing literature has considered all three of these approaches in empirical applications. But empirical evidence as to which approach is best is mixed. Hendry and Hubrich (2011) discuss how performance varies across forecasting or nowcasting situations. As further examples, while Marcellino et al. (2003) and Baffigi et al. (2004) find that aggregating national forecasts delivers better GDP forecasts than forecasts constructed using the aggregate data, in the context of forecasting industrial production Bodo et al. (2000) find the reverse.

In this chapter, we follow Lui and Mitchell (2013b) and unify these alternative approaches for constructing rapid estimates of an aggregate variable. We relate them to a general multivariate framework for forecasting and nowcasting aggregates, and consider various approximations to this multivariate model. In practice, estimation of the multivariate model becomes infeasible as the number of countries, regions or sectors — and the number of indicator variables — increases. These approximations to the infeasible but efficient multivariate model include the use of global VAR models, the aforementioned mixed approach, factor models and Bayesian large-N (shrinkage) VAR models. Many of these models may be implemented using mixed-frequency (unbalanced panel) data, as it is common when producing rapid estimates to exploit indicator variables available at a higher frequency.

The plan of this chapter is as follows.Section 15.2 considers the relationship between aggregate, disaggregate and multivariate models for the construction of rapid estimates of aggregate variables. It explains that, absent parameter estimation error, the multivariate approach is the most efficient method to estimate an aggregate, as it conditions on all available information when forming the rapid estimate. A simple example is also given to show if and how the aggregate, disaggregate and multivariate approaches are equivalent. But, in practice, as the multivariate model is often of too high a dimension to estimate by classical methods, in Section 15.3 we consider approximations which include the use of factor models, global VAR models and the mixed approach of Hendry and Hubrich (2011). Section 15.4 then undertakes a detailed real-time application producing early estimates of Euro Area GDP growth using the aggregate, disaggregate and multivariate approaches. Particular focus is on use of global VAR models to produce the nowcasts, given the novelty and ease-of-use of this approach. Section 15.5 concludes.

Throughout we use ‘country’ to denote the disaggregate of interest, and assume the aggregate of interest involves adding together these countries; in the application below focus is on the Euro Area (EA). But, without loss of generality, these disaggregates could be a region or sector of the economy, for example, instead.

15.2 The relationship between Aggregate, Disaggregate and Multivariate Models

In this section, we introduce the indirect or disaggregate and direct or aggregate approaches; and then show how they can be nested within a multivariate model.

Let $\mathbf{x}_t$ denote a kN-dimensional vector, $\mathbf{x}_t = (\mathbf{x}_{1t},\ \mathbf{x}_{2t}, ...,\ \mathbf{x}_{Nt})\prime$, $(t = 1, ..., T)$. Each $\mathbf{x}_{it}$ is a k-dimensional vector that contains the 'domestic' (disaggregate) variables of country (region or sector) i $(i = 1, \cdots,\ N)$. $\mathbf{x}_{it}$ could simply contain the variable of interest, say domestic GDP growth, $\Delta \log y_{it}$. But it might also contain a $(k-1)$ set of indicator variables for country (region or sector) i, where these indicator variables are chosen on the basis of their timeliness and their informational content. In particular, these indicators maybe published ahead of the variable of interest; and they maybe available at a different (higher) frequency. We denote these $(k-1)$ indicators by $\mathbf{x}_{it}^{Ind}$ and for ease of exposition do not distinguish the frequency of the observations.

The objective is to nowcast the aggregate $\sum_{i=1}^{N} w_i \mathbf{e}'\mathbf{x}_{iT+1}$, where w_i are a set of weights $\left(\sum_{i=1}^{N} w_i = 1\right)$, discussed below, and $\mathbf{e}$ is a binary $(0, 1)$ conformable selection vector which picks out the variable of interest for country i, $\forall i$. For example, it may extract the first element of $\mathbf{x}_{it}$ (say, GDP growth) for each country, i, so that these can then be aggregated using the weights w_i to form aggregate GDP growth.

The indirect (or disaggregate) approach involves first estimating, separately for each country i, a disaggregate model, relating the variable of interest, say GDP growth in country i at period t, to the $(k-1)$ vector of indicators for country i, $(t = 1, ..., T)$, and then aggregating up the nowcasts across i to produce a nowcast for the aggregate:

$$\sum_{i=1}^{N} w_i E\left(\mathbf{e}'\mathbf{x}_{iT+1} \mid \widehat{\mathbf{x}}_{iT+1}^{Ind}, \mathbf{x}_{iT}\right), \tag{15.1}$$

where $\widehat{\mathbf{x}}_{iT+1}^{Ind}$ denotes known and/or forecast values for the indicators $\mathbf{x}_{iT+1}^{Ind}$. Depending on how long after the end of the period T the nowcasts for $(T+1)$ are produced, more of the indicators may, in fact, be known for the whole period $(T+1)$. There is, therefore, then no need to forecast them. For other indicators, partial within-period higher-frequency data might be available and therefore exploited when forming this conditional expectation. Lagged variables, here denoted $\mathbf{x}_{iT}$, are also conditioned on when forming the disaggregate estimate.

The direct (or aggregate) approach involves producing nowcasts directly from an aggregate model. If the disaggregate models are VAR(MA) models, then the aggregate model in $\sum_{i=1}^{N} w_i \mathbf{e}'\mathbf{x}_{it}$ is, in fact, also a VAR(MA) model; see Lütkepohl (1984). Having estimated the aggregate model, the nowcast is computed directly from it as

$$E\left(\sum_{i=1}^{N} w_i \mathbf{e}'\mathbf{x}_{iT+1} \mid \left\{\sum_{i=1}^{N} w_i \mathbf{e}'\widehat{\mathbf{x}}_{iT+1}^{Ind}, \sum_{i=1}^{N} w_i \mathbf{e}'\mathbf{x}_{iT}\right\}\right). \tag{15.2}$$

Neither of these two approaches accommodate dependencies across i; in the disaggregate approach, (15.1), the rapid estimate for country i is conditioned only on indicators for country i. Accommodating dependencies across i might be especially felicitous when nowcasting given that different countries publish their data to different timescales. It might therefore be possible to condition the nowcast on known or partially known, rather than forecasted, disaggregate information. This means for example, as discussed below, that 30 days after the end of the quarter T one can condition the aggregate nowcast on known country-level quarter $(T+1)$ GDP data for those countries, like Belgium and Spain, who currently publish their GDP data within 30 days, as well as Portuguese IP since it is also known for all three months of quarter $(T+1)$.

To accommodate cross-country dependencies when nowcasting, as well as cross-variable dependencies, we

consider the following generic multivariate linear model, where one lag is assumed for notational ease only,

$$\mathbf{x}_t = \Theta \mathbf{x}_{t-1} + \mathbf{u}_t, \text{ or} \tag{15.3}$$

$$\begin{bmatrix} \mathbf{x}_{1t} \\ \mathbf{x}_{2t} \\ \vdots \\ \mathbf{x}_{Nt} \end{bmatrix} = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \cdots & \Theta_{1N} \\ \Theta_{21} & \Theta_{22} & \cdots & \Theta_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Theta_{N1} & \Theta_{N2} & \cdots & \Theta_{NN} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1t-1} \\ \mathbf{x}_{2t-1} \\ \vdots \\ \mathbf{x}_{Nt-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \\ \vdots \\ \mathbf{u}_{Nt} \end{bmatrix}. \tag{15.4}$$

Assume the errors, $\mathbf{u}_t$ follow a multivariate normal distribution, so $\mathbf{u}_t \sim N(0,\ \mathbf{\Sigma})$:

$$\mathbf{\Sigma} = \begin{bmatrix} Cov(\mathbf{u}_{1t},\ \mathbf{u}_{1t}) & Cov(\mathbf{u}_{1t},\ \mathbf{u}_{2t}) & \cdots & Cov(\mathbf{u}_{1t},\ \mathbf{u}_{Nt}) \\ Cov(\mathbf{u}_{2t},\ \mathbf{u}_{1t}) & Cov(\mathbf{u}_{2t},\ \mathbf{u}_{2t}) & \cdots & Cov(\mathbf{u}_{2t},\ \mathbf{u}_{Nt}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\mathbf{u}_{Nt},\ \mathbf{u}_{1t}) & Cov(\mathbf{u}_{Nt},\ \mathbf{u}_{2t}) & \cdots & Cov(\mathbf{u}_{Nt},\ \mathbf{u}_{Nt}) \end{bmatrix}. \tag{15.5}$$

It can be seen that in the generic multivariate model, each partial system for $\mathbf{x}_{it}$ is a VAR(1) model. However, interrelations between these N partial systems arise in two ways: (1) $\Theta_{ij} \neq 0$ and (2) $Cov(\mathbf{u}_{it},\ \mathbf{u}_{jt}) = \mathbf{\Sigma}_{ij} \neq 0$.

A forecast of the aggregate produced by this multivariate system is defined as

$$\sum_{i=1}^{N} w_i E\left(\mathbf{e}'\mathbf{x}_{iT+1} \mid \left\{\widehat{\mathbf{x}}_{iT+1}^{Ind}, \mathbf{x}_{iT}\right\}_{i=1}^{N}\right), \tag{15.6}$$

where w_i denotes the weight assigned to country i. To compute this forecast, which now conditions on all disaggregate information as well as aggregate information, requires estimation of (15.3).

It is well known that classical estimation of (15.3) is only feasible when N is small. When N gets big, estimation of this system involves estimation of a large number of parameters and becomes infeasible. In section 15.3 below we consider various ways of overcoming this curse of dimensionality.

Absent estimation uncertainty, conditioning the aggregate nowcast on all available disaggregate information, via (15.3) and (15.6), is root mean squared error (RMSE) prediction error minimising. The direct, (15.2), and indirect, (15.1), approaches are, in general, inefficient, given that they do not accommodate all of the cross-variable and cross-country dependencies and condition on incomplete information sets. But these methods still deliver unbiased estimates albeit using an incomplete information set.

The indirect approach is a special case of the multivariate model, (15.3), when $\Theta_{ij} = 0$ and $\mathbf{\Sigma}_{ij} = 0$ for all $i \neq j$. Giacomini and Granger (2004) detail the RMSE gain of the multivariate model relative to the direct and indirect approaches. When there is a significant common factor which explains the correlation across i, there can be gains to using disaggregated models provided the common factor is taken into account, which it is not in the underlying disaggregate VAR(1) model, $\mathbf{x}_{it} = \Theta_{ii}\mathbf{x}_{it-1} + \mathbf{u}_{it}$, since the process for the aggregate is dominated by the common factor; cf. Granger (1987). Below, via a simple example, we provide more intuition on when we might expect performance to vary across the direct, indirect and multivariate models. But the conclusion is that the relative gain in efficiency of conditioning the aggregate forecast on all disaggregate components (i.e., of using a multivariate model) is likely large when there is a high degree of heterogeneity in the disaggregate series being aggregated. When the disaggregate series are highly correlated, the relative gain in efficiency is likely low. When the disaggregate series exhibit common variation (e.g., there is a common factor which explains all of their variation) then the direct, indirect and multivariate approaches merge; see Taylor (1978). And, as seen, when there is heterogeneity across countries but $\Theta_{ij} = 0$ and $\mathbf{\Sigma}_{ij} = 0$ for all $i \neq j$, the multivariate approach reduces to the indirect approach.

But the relative performance, in terms of RMSE, of the different approaches is, in practice, given parameter estimation error (and temporal instabilities; cf. Hendry and Hubrich (2011)), an empirical matter.

15.2.1 Equivalence between the aggregate, disaggregate and multivariate approaches: A simple example

Taylor (1978) in his comment on Geweke (1978) uses a simple model to explain temporal and sectoral aggregation of seasonally adjusted data. Although our attention is confined to the three forecasts above, his example can be useful to illustrate the conditions when they are equivalent.

Suppose there are two disaggregate components of the aggregate denoted x_1 and x_2 (for simplicity, we ignore the time subscript). The illustration of Taylor (1978) assumes both series consist of a seasonal and nonseasonal part. We modify this example by simply assuming each series consists of an observed (with superscript A) and unobserved component (with superscript B):

$$x_i = x_i^A + x_i^B, i = 1, 2. \tag{15.7}$$

Further assume the pair of observed and unobserved components, $\left\{x_i^A\right\}_{i=1}^2$ and $\left\{x_i^B\right\}_{i=1}^2$ are uncorrelated, both have zero means and covariance Σ^A and Σ^B respectively. Σ^A and Σ^B represent the variance-covariance matrix between the observed and unobserved components across the two disaggregate series.

Suppose our interest is to obtain the aggregate of the observed component $\mathbf{x} = x_1^A + x_2^A$. Let $\widehat{\mathbf{x}}$ denote a prediction of the aggregate. The optimal prediction of the aggregate is the one that minimises the prediction error

$$E\left[(\mathbf{x} - \widehat{\mathbf{x}})^2 \mid x_1,\ x_2\right].$$

So the optimal prediction should be given by the mean of the aggregate conditional on the disaggregate components x_1 and x_2. That is

$$\widehat{\mathbf{x}}^{(M)} = E\left[\mathbf{x} \mid x_1,\ x_2\right] = E\left[x_1^A + x_2^A \mid x_1,\ x_2\right] = E\left[x_1^A \mid x_1,\ x_2\right] + E\left[x_2^A \mid x_1,\ x_2\right]. \tag{15.8}$$

However the prediction error will not be minimised if the prediction takes the values of

$$\widehat{\mathbf{x}}^{(D)} = E\left[x_1^A \mid x_1\right] + E\left[x_2^A \mid x_2\right] \tag{15.9}$$

or

$$\widehat{\mathbf{x}}^{(A)} = E\left[x_1^A + x_2^A \mid x_1 + x_2\right]. \tag{15.10}$$

We can see that (15.8), (15.9) and (15.10) are analogous to (15.6), (15.1) and (15.2), respectively. That is,

$$\widehat{\mathbf{x}}^{(M)} = \text{multivariate nowcast,} \tag{15.11}$$
$$\widehat{\mathbf{x}}^{(D)} = \text{disaggregate nowcast,} \tag{15.12}$$
$$\widehat{\mathbf{x}}^{(A)} = \text{aggregate nowcast.} \tag{15.13}$$

Taylor (1978) points out that the reason (15.8) is optimal, as compared to (15.9) is that the former is conditional on all available disaggregate components, while the latter is not. Even though the prediction (or nowcast) in (15.1) and (15.9) are the aggregates of the predictions of the disaggregate components, the prediction of the disaggregate component is formed by conditioning on a limited set of information. As we can see from (15.1), nowcasts of the disaggregate of country i are conditional on its own domestic components, similarly as in (15.9). Whereas the nowcast produced for each country, before aggregation, from the multivariate approach (15.6) is conditional on the disaggregate components of all N countries, similarly in (15.8). In other words, $\widehat{\mathbf{x}}^{(M)}$ allows for interdependence among countries and variables, so that the disaggregate information can be fully utilised.

The relative gain in efficiency by conditioning the aggregate forecast on all disaggregate components (in other words, via a multivariate framework) is large when there is a high degree of heterogeneity in the disaggregate series being aggregated. In the case when the disaggregate series are highly correlated, the relative gain in efficiency of such forecast is low. Taylor (1978) further explains the situation when the three predictions are equivalent. This can be appreciated by supposing that $\left\{x_i^A\right\}_{i=1}^2$ and $\left\{x_i^B\right\}_{i=1}^2$ follow a joint normal distribution. Rewrite (15.8), (15.9) and (15.10) as

$$\widehat{\mathbf{x}}^{(M)} = a_1 x_1 + a_2 x_2, \tag{15.14}$$

$$\widehat{\mathbf{x}}^{(D)} = b_1 x_1 + b_2 x_2, \tag{15.15}$$

$$\widehat{\mathbf{x}}^{(A)} = c\left(x_1 + x_2\right). \tag{15.16}$$

When $\Sigma^A = \Sigma^B$ then $a_1 = a_2 = b_1 = b_2 = c$. In this case the three predictions are equivalent. Recall Σ^A and Σ^B represent the two sources of variations in the disaggregate series. $\Sigma^A = \Sigma^B$ implies the disaggregate series exhibit common variation; there is a common factor which explains all of their variation. If the covariance between x_1^A and x_2^A, and the covariance between x_1^B and x_2^B is zero, and there is no homogeneity between the two disaggregates, then the forecast from the multivariate model (in other words, the efficient forecast) is equivalent to the forecast from the disaggregate model and $\widehat{\mathbf{x}}^{(M)} = \widehat{\mathbf{x}}^{(D)}$. Applying this to our models above, this implies $Cov\left(\mathbf{u}_{it},\ \mathbf{u}_{jt}\right) = 0$ in (15.5).

The degree of commonality between the disaggregates relates to the cross-sectional dependence literature, see Pesaran (2006), and also the literature on factor approximations; see Forni et al. (2000) and Stock and Watson (2002b). So when the disaggregate series exhibit a certain degree of heterogeneity in their movement, the multivariate model will be preferred as the different interactions among the disaggregate series can be accounted for. In such a case, we should expect the RMSE of the multivariate model to be less than that of the other approaches.

15.3 Approximations to the multivariate model

Multivariate models therefore benefit, absent parameter estimation error, from minimum RMSE relative to both the aggregate and disaggregate approaches. However, estimation of the multivariate model when N is large is always a challenge, in practice.

The dimensionality problem when N is large can be resolved by shrinking the parameter space and/or reducing the size of the dataset. The former is often achieved by imposing priors and estimating the (large-dimensional) VAR using Bayesian methods (a BVAR); e.g. see Banbura et al. (2010). Giacomini and Granger (2004) restrict the degree of interdependence across i in (15.3) by proposing spatial autocorrelation models. Factor methods are a widespread means of reducing the size of the dataset. Chudik and Pesaran (2011) deal with this curse of dimensionality by effectively shrinking the parameter space in the limit as the number of endogenous variables (N) tends to infinity. This infinite dimensional VAR, as Chudik and Pesaran (2011) explain, can then be arbitrarily well approximated by a set of finite-dimensional small-scale models that can be consistently estimated separately as a global VAR (GVAR) model as proposed in Pesaran et al. (2004). Giannone and Reichlin (2009) discuss the relationship between BVAR and GVAR approaches. In contrast the so-called “mixed approach” by Hendry and Hubrich (2011) places the disaggregate indicators in the aggregate model to produce aggregate forecasts.

Below we review these alternative approximations to the multivariate model. We do not consider in any detail Bayesian solutions to the dimensionality problem which involve shrinkage; for an example see Banbura et al. (2010) who estimate large N Bayesian VAR models.

15.3.1 Factor models

Factor models have been used extensively as an approximation to multivariate models to resolve the problem of dimensionality; see for example Forni et al. (2000), Stock and Watson (2002a) and Stock and Watson

(2002b). Factor models assume the comovement among data series within a large dataset can be summarised by a few unobserved common factors.

Suppose $\mathbf{x}_t$ has a factor structure and can be represented by a factor model

$$\mathbf{x}_t = \mathbf{\Lambda}\mathbf{F}_t + \upsilon_t, \tag{15.17}$$

where $\mathbf{\Lambda}$ is a $kN \times r$ matrix of factor loadings, $\mathbf{F}_t$ is a r-vector of factors capturing the comovement among the variables/indicators in $\mathbf{x}_t$ and υ_t are idiosyncratic disturbances. The common factors can either be static or dynamic, parametrically or non-parametrically estimated.

A prediction of the aggregate using the factor model is then

$$E\left(\sum_{i=1}^{N} w_i \mathbf{e}' \mathbf{x}_{iT+1} \mid \widehat{\mathbf{F}}_{T+1}, \mathbf{F}_T\right), \tag{15.18}$$

where we condition on $\widehat{\mathbf{F}}_{T+1}$ as well as $\mathbf{F}_T$ to capture the fact that in some applications the factors may be known or partially known (forecast, denoted with the $\wedge$) for period $(T+1)$. One could form this forecast or conditional expectation, (15.18), using either one-step or two-step factor estimation methods. The latter is perhaps computationally most convenient; and would involve first using, say, principal components to extract the common factors from the vector $\mathbf{x}_t$. Secondly, these factors could be related to the aggregate via a linear regression which could then be used to compute the forecast; cf. Stock and Watson (2002b). Giannone et al. (2008) offer an alternative where the parameters of the model are estimated consistently through principal components, and the Kalman filter is used to update the estimates of the signal and produce the forecast even with unbalanced panels. Alternatively the factors might be projected not on the aggregate directly but on the disaggregates, and then one could use the factors as in the disaggregate approach.

In contrast to the factor approach the multivariate model can be seen to condition the nowcast for the aggregate not just on the factors but the idiosyncratic terms too. Therefore, the worse the fit of the factor model — in terms of explaining the $\mathbf{x}_t$ — the greater the loss of efficiency of the factor approach relative to the multivariate nowcast, (15.6).

15.3.2 Global VAR model

The Global VAR (GVAR) model developed by Pesaran et al. (2004), Dees et al. (2007) and Pesaran et al. (2009) facilitates estimation of (15.3) by viewing (15.3) as a reduced-form based on N country-specific so-called VARX* models in $\mathbf{x}_{it}$, with international linkages captured by relating $\mathbf{x}_{it}$ to their corresponding foreign variables, $\mathbf{x}^*_{it}$, defined below. Estimation then proceeds in two steps. First, estimate by OLS N VARX* models for each country to obtain the parameter matrices in the country-specific models. The GVAR in the endogenous variables $\mathbf{x}_{it}$ is then solved for simultaneously by using a "link' matrix that contains the predetermined weight for each country and the parameter matrices in the N VARX* models.

Specifically, consider a VARX* model for country i $(i = 1, \cdots, N)$, again assumed to be first-order and ignoring deterministic terms,

$$\mathbf{x}_{it} = \Phi_{i0}\mathbf{x}_{it-1} + \Lambda_{i0}\mathbf{x}^*_{it} + \Lambda_{i1}\mathbf{x}^*_{it-1} + \mathbf{u}_{it} \tag{15.19}$$

where $\mathbf{x}^*_{it}$ is a $k^* \times 1$ vector containing the foreign variables. These foreign variables can be seen to capture the effect of common factors; cf. Pesaran (2006). $\mathbf{x}^*_{it}$ vary by country i and are defined as

$$\mathbf{x}^*_{it} = \sum_{j=1}^{N} w_{ij}\mathbf{x}_{jt}; w_{ii} = 0 \tag{15.20}$$

with the predetermined weights w_{ij} capturing the importance of country j for country i. Importance can be measured in different ways, depending on the application. In the application to GDP growth below we use GDP shares, namely the size of country j's GDP relative to the sum of GDP in the other $N-1$ (excluding country i) countries. In principle one might let these weights vary over time. There is in fact limited time variation in GDP shares over our sample; and we estimate the weights as the average over the period 1998q1 to 2003q1.

When $\mathbf{x}_{it}^{*}$ are not clearly weakly exogenous then consistent estimation of (15.19) is best approached by first estimating, for each country i, a $(k+k^{*})\times 1$ dimensional VAR in $\mathbf{z}_{it}=(\mathbf{x}_{it}^{\prime},\mathbf{x}_{it}^{*\prime})^{\prime}$ and then, at a second step, using these parameter estimates to solve for the parameters of the conditional model, as seen in (15.19). Given a nowcasting focus and desire to use statistical rather than economic models, the VAR might be estimated using stationary data (i.e. data are differenced until stationary) and any long-run, cointegrating, restrictions can be ignored.

Since the domestic variables $\mathbf{x}_{it}$ are related contemporaneously to the foreign variables $\mathbf{x}_{it}^{*}$ the country specific VAR models (15.19) need to be solved simultaneously for all of the endogenous domestic variables $\mathbf{x}_{it}$ $(i=1,...,N)$. This is achieved by first re-writing equation (15.19) as

$$\mathbf{A}_{i}\mathbf{z}_{it}=\mathbf{a}_{i0}+\mathbf{B}_{i}\mathbf{z}_{it-1}+\mathbf{u}_{it} \tag{15.21}$$

where $\mathbf{A}_{i}=(\mathbf{I}_{k},-\mathbf{\Lambda}_{i0})$, and $\mathbf{B}_{i}=(\mathbf{\Phi}_{i0},\mathbf{\Lambda}_{i1})$. Then, since the domestic variables can be written as

$$\mathbf{z}_{it}=\mathbf{W}_{i}\mathbf{x}_{t} \tag{15.22}$$

where $\mathbf{W}_{i}$ is a $(k+k^{*})\times kN$ link matrix defined by the (fixed, not estimated) GDP share weights.

Using (15.22) in (15.21), we have

$$\mathbf{A}_{i}\mathbf{W}_{i}\mathbf{x}_{t}=\mathbf{B}_{i}\mathbf{W}_{i}\mathbf{x}_{t-1}+\mathbf{u}_{it} \tag{15.23}$$

The GVAR model is thus derived by stacking up all N country-specific VARX* models to yield

$$\begin{aligned}\mathbf{G}\mathbf{x}_{t} &= \mathbf{H}\mathbf{x}_{t-1}+\mathbf{u}_{t}\\ \mathbf{x}_{t} &= \mathbf{G}^{-1}\mathbf{H}\mathbf{x}_{t-1}+\mathbf{G}^{-1}\mathbf{u}_{t}\end{aligned} \tag{15.24}$$

where

$$\mathbf{G}=\begin{bmatrix}\mathbf{A}_{1}\mathbf{W}_{1}\\ \vdots\\ \mathbf{A}_{N}\mathbf{W}_{N}\end{bmatrix},\qquad \mathbf{H}=\begin{bmatrix}\mathbf{B}_{1}\mathbf{W}_{1}\\ \vdots\\ \mathbf{B}_{N}\mathbf{W}_{N}\end{bmatrix},\qquad \mathbf{u}_{t}=\begin{bmatrix}\mathbf{u}_{1t}\\ \vdots\\ \mathbf{u}_{Nt}\end{bmatrix}, \tag{15.25}$$

and $\mathbf{u}_{t}$ has the unrestricted covariance matrix Σ, which can be estimated using each country's k-vector of residuals $\widehat{\mathbf{u}}_{it}$.

Having estimated the N VARX* models, and having solved for the GVAR, nowcasts are computed from (15.24) by computing the expectation of $\mathbf{x}_{T+1}$ (specifically, the GDP growth elements of $\mathbf{x}_{T+1}$) conditional on $\widehat{\mathbf{x}}_{iT+1}^{Ind}$ and lags of $\mathbf{x}_{T}$. Nowcasts for the aggregate, EA GDP growth, are then obtained by taking the weighted average of these conditional expectations for GDP growth across the individual countries. See Lui and Mitchell (2013a) and Lui and Mitchell (2013b) for more on the use of the GVAR to produce nowcasts.

15.3.3 Mixed model: using disaggregate information in the aggregate model

We have considered above the factor and GVAR models as approximations to the multivariate model. Both aim to reduce the dimension of the disaggregate information set to make estimation feasible. Therefore,

both methods can be seen to nest the disaggregate approach, with the former summarising the disaggregate information via the common factors, the latter directly estimating disaggregate models then using the output to construct the multivariate (global) model. But in neither method is aggregate information both used as the dependent and explanatory variable too. This is in contrast to the so-called 'mixed approach' of Hendry and Hubrich (2011) which directly seeks to condition on the same information set as in the multivariate approach, (15.2).

Following Hendry and Hubrich (2011) re-consider the generic multivariate model (15.3) in the disaggregates:

$$\mathbf{x}_t = \Theta \mathbf{x}_{t-1} + \mathbf{u}_t. \tag{15.26}$$

Pre-multiplying (15.26) by the weight vector $\mathbf{w} = \mathbf{w}^* \otimes \mathbf{e}$, where $\mathbf{w}^* = (w_1, ..., w_N)$ and $\mathbf{e}$ is again the k-vector (0, 1) selection vector to extract the variable of interest:

$$\mathbf{w}\prime\mathbf{x}_t = \mathbf{w}\prime\Theta\mathbf{x}_{t-1} + \mathbf{w}\prime\mathbf{u}_t \tag{15.27}$$

$$\mathbf{w}\prime\mathbf{x}_t = \kappa\mathbf{w}\prime\mathbf{x}_{t-1} + (\mathbf{w}\prime\Theta - \kappa\mathbf{w}\prime)\,\mathbf{x}_{t-1} + \mathbf{w}\prime\mathbf{u}_t \tag{15.28}$$

we see that the aggregate $\mathbf{w}\prime\mathbf{x}_t$, in the application EA GDP growth, depends on the lagged aggregate $\mathbf{w}\prime\mathbf{x}_{t-1}$ and the vector of lagged disaggregates $\mathbf{x}_{t-1}$.On this basis Hendry and Hubrich (2011) explain that the aggregate model can be improved systematically by including additional disaggregate components to the extent that the parameters of the disaggregate components in (15.28) are constants, i.e. $(\mathbf{w}\prime\Theta - \kappa\mathbf{w}\prime) \neq \mathbf{0}$, and the disaggregate components are significant in explaining the aggregate.

The optimal prediction of the aggregate is then given by its conditional mean

$$E\left(\mathbf{w}\prime\mathbf{x}_{T+1} \mid \mathbf{w}\prime\mathbf{x}_T,\ \{\mathbf{x}_{iT}\}_{i=1}^N\right), \tag{15.29}$$

where again we could condition on known or forecast within-period information too, i.e. $\{\widehat{\mathbf{x}}_{iT+1}^{Ind}\}_{i=1}^N$, although we suppress this here to keep notation simple.

We can see that the nowcast (15.29) is based on a richer information set than (15.2); like the multivariate approach it efficiently conditions on all available aggregate and disaggregate information. Nothing is lost by predicting the aggregate directly, rather than predicting the disaggregates and then aggregating. See also Castle and Hendry (2010).

Hendry and Hubrich (2011) explain that differences in the performance of different forecasting approaches can be explained by: (i) changes in collinearity between the disaggregate constituents that in turn affect the bias-variance trade-off in forecast model selection; (ii) changes in the proportion of the variance of the disaggregate constituents that is captured by the common factor; (iii) changes in the unconditional moments of the aggregate being forecasted. When the aggregate and the disaggregate series exhibit a considerable amount of variability, use of disaggregate variables to forecast the aggregate can improve forecast accuracy. Hendry and Hubrich (2011) also undertake an empirical analysis comparing alternative means of forecasting an aggregate. The findings of their empirical application to forecast Euro Area inflation provide weak evidence indicating that disaggregate information helps improve forecast accuracy for the aggregate. But for US inflation their results are contrasting; this again emphasises that it varies across applications whether, in practice, it helps to condition on disaggregate information when forecasting an aggregate.

In practice estimation of the mixed model, (15.28), may also involve estimation of many parameters, given that $\mathbf{x}_{t-1}$ is a right-hand-side variable in the regression against the aggregate $\mathbf{w}\prime\mathbf{x}_t$. And $\mathbf{x}_{t-1}$ can quickly become large dimensional, and therefore (15.28) would involve estimation of many parameters relating these disaggregates to the aggregate. Model selection or model combination may therefore be used. Model selection would involve a search for the subset of disaggregate indicators from $\mathbf{x}_{t-1}$ that best explain the aggregate. Model

combination would involve estimating many aggregate models which differ in the (smaller) set of disaggregate indicator employed: a combination then provides one means of pooling information across the different elements of $\mathbf{x}_{t-1}$. One attractive strategy is to follow Castle and Hendry (2010) and Castle et al. (2013) and use automatic model selection methods (*Autometrics*) to estimate (15.28) directly when there are more variables than observations, i.e $Nk > T$, and when these variables are perhaps perfectly collinear. Bayesian (shrinkage) methods offer an alternative when $Nk > T$.

Castle et al. (2013) consider forecasting with factors, variables and both, modelling in-sample using model selection methods so all factors (principal components) and variables can be included jointly. This therefore amounts, in our context, to conditioning the aggregate nowcast not just on $\left(\mathbf{w}\prime\mathbf{x}_T,\ \{\mathbf{x}_{iT}\}_{i=1}^{N}\right)$ but also on $\left(\widehat{\mathbf{F}}_{T+1}, \mathbf{F}_T\right)$.

15.4 Application: Nowcasting Euro area GDP growth

We empirically compare the performance of direct and indirect nowcasts of quarterly EA GDP growth with those constructed using a Global VAR model (strictly, a European VAR model), as proposed by Pesaran et al. (2004), which allows for both cross-country and cross-variable dependencies. The GVAR model, as we have explained, offers a computationally convenient and feasible means of producing nowcasts of aggregates like EA GDP growth from large disaggregate datasets, via a multivariate model which can capture any dependencies.

Our particular focus as in Lui and Mitchell (2013b) is the construction of nowcasts for EA GDP growth 30 days after the end of the quarter, which is 15 days earlier than Eurostat's own *Flash* estimate and coincident with when *advance* US GDP and *preliminary* UK GDP data are first published. But to quantify the putative trade-off between the timeliness and accuracy of the nowcasts we consider how nowcasts of quarterly GDP can be constructed as within-quarter (monthly) information on indicator variables accrues. Due to differing publication lags for data across the EA countries the amount of within-quarter information available at a given point in time (after the end of the quarter of interest) varies across countries. Focus on the EA(12) (i.e., $N = 12$) means the twelve countries comprising the aggregate are Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain.

Here we consider forecasting quarterly GDP growth using quarterly regression and quarterly GVAR models, with a small number of indicator variables, using 'bridge' models to fill in missing quarterly data. Bridging involves linking monthly indicator data, typically released early in the quarter, with quarterly data like GDP; e.g., see Salazar and Weale (1999), Baffigi et al. (2004) and Diron (2008). In effect, a two-equation system is used to nowcast, with the second equation comprising the forecasting model for the monthly variable. The errors between the two equations, at the underlying monthly frequency, are assumed orthogonal so that the equations are estimated separately. In common with much previous work, see Diron (2008), we experimented with simple monthly AR models to forecast any missing monthly data. This included using both an AR(1) model to forecast the missing month(s) and a random walk type model in first-differences where the missing month(s) is (are) the mean of the available within-quarter data. In fact, we found some gains to use of the random walk forecast and therefore focus on its use in the application below. An alternative to bridging is to nowcast and forecast monthly GDP using mixed-frequency regression, VAR or factor-based methods, which impose an aggregation constraint so that monthly GDP is consistent with the published quarterly values (e.g., see Mariano and Murasawa (2003), Mitchell et al. (2005), Angelini et al. (2010), Angelini et al. (2011) and Banbura and Rünstler (2011)). Bayesian estimation methods, exploiting the Kalman filter and 'skipping' missing observations, and MIDAS models (e.g. see Kuzin et al. (2011)) also offer alternative means of linking mixed-frequency data.[1]

[1] We did experiment in the application below with mixed-frequency VAR models. However, we found only marginal improvements in nowcast accuracy at the country-level when mixed-frequency VAR rather than quarterly (bridge) VAR models were used.

Table 15.1: Within-quarter data availability for the EA and EA countries, at t+0, t+15 and t+30 days after the end of quarter t. m1, m2 and m3 denote the first, second and third month in quarter, t

	t+0 days	t+15 days	t+30 days
IP	m1 for all EA countries & EA m1, m2 for Portugal	m1, m2 for all EA countries (excluding Austria & Belgium) and EA. Only m1 for Austria & Belgium	m1, m2 all EA countries & EA m1, m2, m3 for Portugal
Survey	m1, m2, m3 all EA countries & EA	m1, m2, m3 all EA countries & EA	m1, m2, m3 all EA countries & EA
GDP	t-1 for all EA countries & EA excluding Luxembourg t-2 for Luxembourg	t-1 for all EA countries & EA	t-1 for all EA countries & EA t for Belgium and Spain

15.4.1 Aggregate and disaggregate indicator variables

We focus on a parsimonious, but key, set of indicator variables directly related to the target variable, GDP growth. Specifically, as a 'hard' indicator for GDP growth we consider (aggregate and disaggregate, i.e. country-level) industrial production (IP) data. As a component of GDP, but available on a monthly basis, we should expect movements in industrial production to be informative about GDP growth. As additional ('soft") indicators we consider qualitative survey data, again at the aggregate and disaggregate levels. A perceived attraction of these surveys is that they are forward, as well as backward, looking. From the surveys we use the Economic Sentiment Indicator (ESI).[2] The ESI, published by the European Commission, is a widely used composite indicator. The ESI combines various information from qualitative business tendency surveys, including expectations questions, into a single cyclical confidence indicator. We do not consider indicator variables like interest rates since, whatever their influence on demand, we believe that they are not so suitable for the production of (pre-) first estimates of data (as opposed to forecasts) since they do not have such clear links to the data they are supposed to represent. Nevertheless, the (G)VAR approach could be readily extended to consider a wider range of indicator variables, although factor-augmented VAR models might be advisable to prevent the dimension of the GVAR becoming unwieldy. However, limited experimentation with alternative indicator variables, such as retail trade, indicated no clear gains to extending the set of indicators. These additional indicators did not appear to add information relative to the IP and survey data, especially since these data are exploited at the disaggregate level. Previous work has found nowcast accuracy to improve dramatically on receipt of within-quarter IP and survey data (e.g. see Giannone et al. (2008), Banbura et al. (2012) and Mazzi et al. (2013)); with studies finding survey data to deliver gains early in the quarter and the IP data becoming dominant when published later in the quarter.

Publication lags for the indicators

As alluded to above, publication lags vary across variables, countries and indeed time. But there is a pattern to when data are published, with hard data generally published around the middle of the month and soft data published at the end of the month. This pattern can be exploited to produce nowcasts at selected horizons after the end of the quarter of interest. Specifically, we produce nowcasts for quarter t, t+0, t+15 and t+30 days after the end of quarter t. Table 15.1 summarises data availability, across indicators and countries, at these three horizons.[3] Eurostat produce the first official estimate for GDP growth, their *Flash* estimate, at t+45 days.

[2] ESI data are not available for Ireland. Instead we use DG-ECFIN's Consumer Confidence Indicator for Ireland. But because this series was discontinued in April 2008 we extend it using the aggregate (EA) data.

[3] Since publication lags have varied over time, as release calendars are revised, we base our analysis on the 2012/3 release calendars. This assumption primarily affects how we treat the Spanish GDP data, since Spain began producing an earlier GDP estimate, at t+30 days, only recently. When missing the vintage data for this earlier estimate we treat the first available release as the t+30 estimate. Comparison between the two estimates, when available, suggests that this assumption is not a strong one. Nevertheless, below we do summarise results when the Spanish data are not assumed known at t+30 days.

Table 15.1 shows how within-quarter data accumulate across the EA. Since the survey data are known for all months in the quarter even at t+0 days, it is how the data accrue on IP and GDP that matters. Table 15.1 shows that even at t+0 days the first month of within-quarter IP data is known; and in Portugal the second month is also known. By t+15 days we know the second month's IP data for most of the EA. This means only the third month of the quarter (for industrial production) needs to be forecast (in this application we do so using a monthly random walk model, i.e. we "bridge"). Inaccurate forecasts of one month may nevertheless deliver reasonably accurate projections of the quarter to which they belong. But for Portugal at t+30 days there is no need to forecast since national accounts' data are available. Table 15.1 also shows how by t+30 days full-quarter GDP data are (currently) available for Belgium and Spain. The GVAR therefore conditions on these data both when nowcasting GDP for the other countries and when aggregating the twelve disaggregate (country-level) GDP nowcasts. The indirect approach uses the hard data for Belgium and Spain only when aggregating, but not when nowcasting GDP in the other countries. The majority of the other countries in the EA produce their GDP estimates more slowly, after about 45 days, in line with Eurostat's *Flash* estimate for EA GDP. Therefore their GDP data need to be nowcast at t+30 days. Luxembourg's GDP data for quarter t-1 also need to be forecast at t+0 days, given the extra lag in the publication of their data.

15.4.2 Real-time simulation exercise

We compare the accuracy of nowcasts of Euro-area GDP growth 0, 15 and 30 days after the end of the quarter, in recursive out-of-sample experiments using real-time data. Specifically, we use the real-time data triangles for real GDP and industrial production available from Eurostat's real-time EuroIND database for the EA(12) aggregates as well as the twelve countries.[4] This is a unique data source made available by Eurostat and means that in contrast to the majority of applied studies studying the Euro-area, we exploit real-time (aggregate and disaggregate) data. This means that our out-of-sample simulations are genuinely, rather than 'pseudo', real-time. Data availability across the EA means that we use only successive first-release GDP estimates, and do not exploit revised (within-quarter) GDP data. The qualitative survey data are not revised (in a significant manner at least). Models are estimated on data vintages back to 2003q1 with, for a given data vintage (i.e., column of the real time triangle) data back to 1985q1. For several countries historical data were not available prior to 1991q1 and backcasted data were used.[5] Since our focus is on nowcasting accuracy from 2003q2 these historical data were generated to increase the possible size of in-sample estimation windows. Seasonally adjusted data are used.

It is important to use real-time data, namely data available to the forecaster at the time they actually made their forecast, rather than the latest release from Eurostat, given that data are revised. Use of the latest vintage of data may give a misleading impression of the accuracy of a given forecasting model/strategy, since in reality the forecaster used an earlier vintage of the data to make their forecasts. We therefore use the latest vintage of data available to the forecaster when they made their forecast.

The nowcasts are evaluated by defining the 'outturn' as both the first (*Flash*) GDP growth estimate from Eurostat and the latest or 'final' (as of the timing of writing) vintage, which contains GDP data up to 2011q4. The exercise could be repeated for different definitions of the outturn, say the second or third release. But as our primary interest is in accelerating delivery of national accounts data, the first estimate does appear to be the obvious benchmark.

[4] The aggregated real GDP data for the EA12 is computed as the sum (in millions of Euros) of the disaggregate GDP series of the 12 individual countries. These data are virtually indistinguishable from aggregate (EA12) GDP data which can be downloaded directly from Eurostat.

[5] Specifically national GDP data back to 1985q1 were not available for Austria, Germany, Ireland, Luxembourg and Portugal. National data were backcasted, in growth rates, using as an indicator data from that country most highly correlated (from 1991q1) with the country of interest. Experimentation with alternative backcasting strategies led to similar nowcasting results. This is perhaps not surprising, given these backcasted data are historical data (used for estimation) not used for evaluation.

Table 15.2: RMSE of competing nowcasts of EA GDP growth (quarterly growth rates) against Eurostat's first (*Flash*) and "final' estimate, 2003q2-2011q4

	First			"Final"		
	t+0	t+15	t+30	t+0	t+15	t+30
Direct	0.53	0.48	0.48	0.56	0.49	0.49
Indirect	0.40	0.41	0.38	0.43	0.44	0.41
GVAR	0.55	0.49	0.34	0.58	0.53	0.37
Random Walk	0.62	0.62	0.62	0.67	0.67	0.67

15.4.3 Nowcasting results

For the direct, indirect and GVAR models nowcasts of GDP growth are constructed from 3-variable quarterly VAR models (i.e. $k = 3$ and $\mathbf{x}_{it}$ is a 3×1 vector and comprises GDP growth, IP growth and the survey data). In fact, 6-variable unconditional VARs in $\mathbf{x}_{it}$ and their foreign counterparts $\mathbf{x}^*_{it}$ are estimated to accommodate the endogeneity of these foreign variables. There is a question about whether the survey data should be considered in levels or first differences. Results below are based on first differences, but are little affected by the transformation. We focus in the application below on first-order VAR models; as described above, when nowcasting, we have a preference for parsimonious dynamics. But, in any case, consideration of higher-order VARs did not deliver more accurate nowcasts in this application. When monthly indicator data are not available, see Table 15.1, the missing months in the quarter are forecast using a first-differenced random walk type model, whereby the missing months are set equal to the mean of the known within-quarter data. We did also experiment with using recursively estimated AR(1) models to forecast the missing month(s) in the quarter; in almost all cases this resulted in less accurate nowcasts. First-differenced random walk type models are known to offer some protection against structural breaks; see Clements and Hendry (1999).

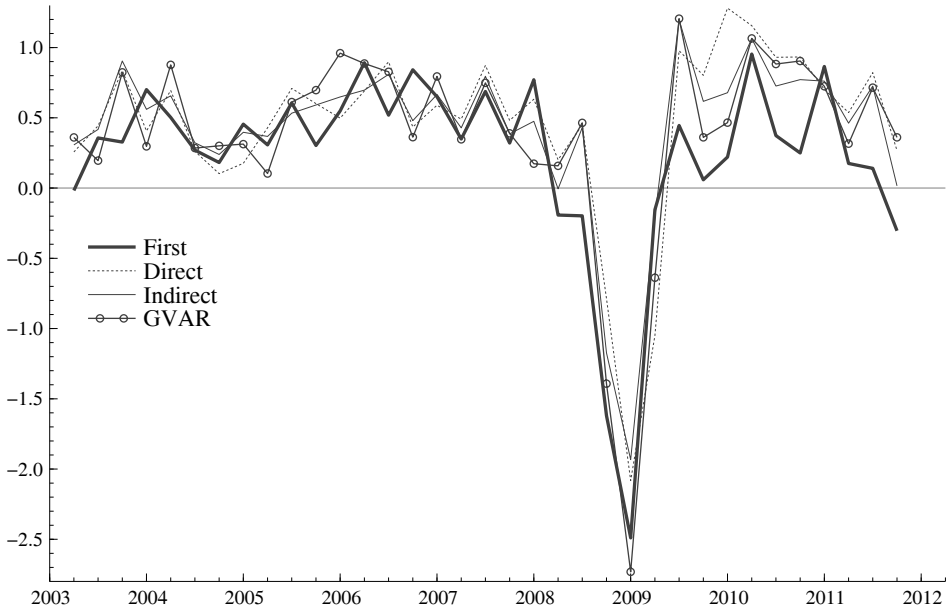
The nowcasts for GDP growth are computed recursively from 2003q2-2011q4 using the direct, indirect and GVAR models. Evaluation of the nowcasts starts in 2003q2, as this is when Eurostat first published its own *Flash* estimate (at t+45 days). For each recursion (data vintage) these models are estimated on data back to 1985q1 (1985q2 in growth rates). This means that an increasing window of data is used for model estimation. Longer and increasing estimation windows appear to help the GVAR, as we discuss further in Section 15.4.3 below.

Table 15.2 summarises, across the three nowcasting horizons, the accuracy of the competing nowcasts by reporting their Root Mean Squared Error (RMSE) statistics against both the first (*Flash*) estimate and the latest (or 'final') estimate. As a benchmark we consider a random walk forecast (results using a recursively estimated AR model are similar). Before looking in more detail at the results at t+30 days, our particular focus given this is when the UK and US currently produce their early GDP estimates, we summarise findings across the three nowcasting horizons.

Across the different nowcasting horizons all three nowcasting models deliver more accurate GDP estimates than the benchmark random walk forecast. Reassuringly, conditioning on within-quarter information improves accuracy. Accuracy also tends to increase as within-quarter information accumulates. But this improvement is most marked for the GVAR. The performance of the indirect approach, itself a restricted GVAR, is fairly constant across the three horizons, declining less steeply than the Direct and GVAR models. Indeed up until t+30 days the nowcasts from the Indirect approach are the most accurate.[6] It is with, at t+30 days, the arrival of full-quarter data for some of the disaggregates (countries) that the GVAR becomes competitive, delivering lower RMSE statistics than the commonly used Direct approach to nowcasting an aggregate and, albeit less so, the Indirect approach. Recall the GVAR efficiently conditions on available disaggregate data. Since the

[6] We prefer not to stress the use of Diebold and Mariano (1995) and Giacomini White (2006) type tests to test the statistical significance of these forecast accuracy differences, given our relatively short evaluation period and the large size of our models with many estimated parameters. But the superiority of the GVAR in Table 15.1 does not appear to amount to statistically significant differences when generic forecast comparison tests are computed over the evaluation period, 2003q2-2011q4, as a whole. We return to this issue below.

Figure 15.1: Competing nowcasts at t+30 days of Eurostat's First (*Flash*) GDP growth estimate published at 45 days



Direct approach does not exploit in any form the new disaggregate data available at t+30 days, its performance is the same as at t+15 days.

To illustrate the benefits of conditioning the GVAR nowcasts (of the aggregate) on country-level GDP data, when available for quarter t, we note that if we conditioned only on Belgian GDP (and not Spanish GDP as in Table 15.2) then the RMSE statistics at t+30 days rise to 0.41 and 0.45 (against the first and final release), respectively. In turn, if the German GDP were published at 30, rather than 45, days after the end of the quarter, and we conditioned on them in the GVAR, the RMSE statistics for the GVAR in Table 15.1 would fall to 0.21 and 0.26, respectively.

To give a clearer impression of relative performance at t+30 days, Figure 15.1 plots the competing nowcasts against Eurostat's first (*Flash*) estimate. This Figure 15.1 shows that, in fact, the GVAR picked up the severity of the descent into recession in 2008 better than the other nowcasting approaches. Of the nowcasting methods, the Direct approach, in particular, detected the downturn only at a lag. But the GVAR did over-estimate slightly both the depth of the recession and the magnitude of the subsequent bounceback; the GVAR nowcasts appear more volatile than their competitors, especially prior to the 2008-9 recession.

To help indicate the quality of these EA nowcasts we compare them with vintage US *advance* estimates (available at t+30 days). These are available from The Federal Reserve Bank of Philadelphia's real-time dataset for macroeconomists. These advance estimates are computed by the BEA using a different methodology, as described in Landefeld et al. (2008). In particular, the BEA have access to and exploit official quantitative surveys, which underlie all official GDP estimates, and at t+30 days have full-quarter information for about 45% of the advance estimate. This means that the BEA have to rely far less on nowcasting/forecasting models using indicators than we do for the EA. Therefore, it is no surprise that at t+30 days the EA nowcasts seen in Table 15.2 are less accurate. Over the 2003q2-2011q4 evaluation period the US advance estimate has a RMSE statistic of 0.19 against their t+120 estimate (a comparable t+45 day estimate is unavailable). We might expect accuracy to be even higher against an earlier estimate. But against the 'final' (again as of time of writing) estimate, the US advance estimate has a RMSE statistic of 0.38. This is comparable to the EA RMSE statistics seen in Table 15.2. This appears to reflect greater revisions to US than EA estimates between t+120 days and the 'final' estimate.

Structural instabilities: robustness to the estimation window

To examine the structural stability of the three nowcasting models, and the robustness of the results in Table 15.2 which are based on estimation over a full-sample (recursive, expanding) window, we use two alternative strategies to select the estimation window.

First, following Pesaran and Timmermann (2007), we combine nowcasts based on the same model estimated over a range of different estimation windows. This allows parameter estimates to vary over time, but does not rely on knowledge of when any breaks might have occurred. Pesaran et al. (2009) found this 'Ave' strategy to be an effective means of improving forecast accuracy from GVAR models, which are liable to suffer from considerable parameter estimation uncertainty. In part reflecting data availability, the precise Ave strategy we adopt is as follows. When nowcasting 2003q2 we consider estimation windows starting in 1985q2, then in 1985q3 and so on until only data from 1993q2 are used. This means the minimum length of the estimation window is 10 years. Then, as we move through the evaluation period, we add one extra estimation window at each recursion, as we keep the minimum estimation window at 10 years.

Secondly, we produce nowcasts from the three models estimated over rolling estimation windows. Rolling windows also offer some protection against structural instabilities. Again reflecting data availability, we consider lengths for the rolling window in the range 10 years (the shortest window) to 18 years (the longest window). We present results for the shortest and longest windows. (In fact, some data-mining reveals that RMSE estimates were minimised for each of the three models either at the shortest or longest window, with RMSE estimates either increasing or decreasing as the window was lengthened.)

Table 15.3 summarises the RMSE results for these two estimation strategies. Comparison with Table 15.2 shows gains to averaging (Ave) in all but one case. Ave delivers lower RMSE statistics than the expanding window at all horizons and across all three models, except for the GVAR at t+30 days. Only in this case does Ave raise the RMSE (against the First estimate), from 0.34 to 0.38. Given the improved performance of the Indirect approach under Ave this means that at t+30 days GVAR, on average across the evaluation period, is no longer the preferred model. But recall that even relative to the expanding estimation window Indirect model seen in Table 15.1 the gains to GVAR were modest, when evaluated over the evaluation period as a whole. Of interest, as seen in Figure 2, is the more obvious advantage GVAR offers over the recessionary period. The dashed line in Figure 2 indicates that the GVAR continues to confer benefits over the recession even relative

Table 15.3: RMSE of competing nowcasts of EA GDP growth (quarterly growth rates) against Eurostat's first (Flash) and "final" estimate, 2003q2-2011q4, for averaging (Ave) and rolling window estimation strategies

	First			"Final"		
	t+0	t+15	t+30	t+0	t+15	t+30
Direct: Ave	0.49	0.45	0.45	0.53	0.47	0.47
Indirect: Ave	0.38	0.36	0.34	0.42	0.40	0.37
GVAR: Ave	0.52	0.44	0.38	0.55	0.46	0.38
Direct: rolling (short)	0.47	0.44	0.44	0.50	0.46	0.46
Indirect: rolling (short)	0.39	0.34	0.32	0.43	0.38	0.35
GVAR: rolling (short)	0.55	0.57	0.53	0.58	0.57	0.55
Direct: rolling (long)	0.51	0.46	0.46	0.54	0.48	0.48
Indirect: rolling (long)	0.39	0.36	0.34	0.43	0.39	0.37
GVAR: rolling (long)	0.62	0.50	0.41	0.63	0.50	0.39

to the more competitive combined (Ave) Indirect model. Although, if the Fluctuation test were computed the gains no longer appear as statistically significant.

Turning to the rolling results reported in Table 15.3 we can understand why averaging across estimation windows does not help the GVAR, at t+30 days, but otherwise does improve nowcast accuracy. Across the three horizons we see modest gains in accuracy for Direct and Indirect if the shorter rolling window is used, rather than the longer window. In turn, the short rolling window delivers lower RMSE estimates than the expanding estimation window used in Table 15.2. This shows that combination can indeed help. But the evidence is more mixed for GVAR. At t+0 days a short rolling window does deliver some gains. But, as within-quarter information accrues, and it becomes more important to estimate the parameters in the GVAR reliably, the GVAR benefits from a longer estimation window. As a result, at t+15 and t+30 days averaging and rolling windows deliver higher RMSE statistics for GVAR than in Table 15.2 when an expanding estimation window was used. The GVAR is more sensitive to the length of the estimation window that both Direct and Indirect.

That there is considerable parameter estimation uncertainty when estimating the GVAR is not surprising in this application, which is characterised by not just a relatively small T sample, but by 'structural instabilities' reflecting the sharp recession in the aftermath of the global financial crisis beginning in 2007.

15.5 Conclusion

This chapter considers the construction of rapid estimates for aggregate variables. It contrasts "direct aggregate" and "indirect disaggregate" modelling approaches. The aggregate forecast/nowcast conditions only on aggregate indicators. Whereas the disaggregate forecast/nowcast conditions on disaggregate components before they are aggregated to form a prediction for the aggregate. However, the latter are still based on a limited information set compared to forecasts/nowcasts produced by a multivariate or mixed model in which not only all available disaggregate components are utilised, but the cross-country, cross-variable dependencies are also accommodated. We outline the conditions under which forecasts/nowcasts produced by this multivariate framework, that fully exploit all available disaggregate components, are equivalent to the aggregate and disaggregate forecasts/nowcasts. We also explore different multivariate modelling approaches, such as factor models, mixed models, global VAR models and Bayesian large-N VARs; which can be seen as approximations to the multivariate model.

We then undertake a detailed set of out-of-sample simulations, using real-time vintage data, to investigate the empirical performance of the aggregate and disaggregate approaches when nowcasting Euro Area GDP growth at different timescales. We contrast performance with a global VAR model, which is designed to condition the nowcast for the EA aggregate on a wider information set. Absent parameter estimation error, the global VAR model delivers minimum RMSE nowcasts. But, in practice, early in the quarter, at t+0 and

t+15 days, there are substantial advantages to using the indirect approach to nowcasting the aggregate. The indirect approach, itself a restricted GVAR that ignores cross-country dependencies, allows for heterogeneity across the EA countries.

But at t+30 days, when more within-quarter data are available, the proposed GVAR nowcasts become more competitive than both the direct and indirect approaches, specifically over the recession. This result is robust to estimation strategies which seek to accommodate structural instabilities. The sensitivity of the different models' relative performances reminds us that parameter estimation error reduces the accuracy of the multivariate model's nowcasts, consistent with results in Lütkepohl (1984) and Giacomini and Granger (2004). Consideration of a wider range of indicators, possibly in conjunction with a factor model or using the mixed-approach of Hendry and Hubrich (2011), offers the potential of further improving nowcast accuracy. Similarly in the face of instabilities and unknown structural breaks it is important to seek to render the nowcasts more robust; cf. Castle et al. (2013).

Bibliography

Angelini, E., Banbura, M.and 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, Vol. 2010, No 1.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L. and Rünstler, G. (2011), 'Short-term forecasts of euro area GDP growth', Econometrics Journal, Vol. 14, No 1, pp. C25–C44.

Baffigi, A., Golinelli, R. and Parigi, G. (2004), 'Bridge models to forecast the euro area GDP', International Journal of Forecasting, No 20, pp. 447–460.

Banbura, Marta 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, Vol. 27, No 2, pp. 333–346.

Banbura, M., Giannone, D. and Reichlin, L. (2010), 'Large Bayesian Vector Autoregressions', Journal of Applied Econometrics, Vol. 25, pp. 71–92.

Banbura, M. and Giannone, D. and Reichlin, L. (2012), 'Nowcasting', Oxford Handbook on Economic Forecasting, eds Michael P. Clements and David F. Hendry, Oxford University Press.

Bodo, G., Golinelli, R. and Parigi, G. (2000), 'Forecasting Industrial Production in the Euro Area Economy', Empirical Economics, Vol. 25, pp. 541–561.

Castle, J. L. and Hendry, D. F. (2010), 'Nowcasting from disaggregates in the face of location shifts', Journal of Forecasting, Vol. 29, No 1-2, pp. 200–214.

Castle, J. L., Clements, M. P. and Hendry, D. F. (2013), 'Forecasting by factors', by variables, by both or neither?, Journal of Econometrics, Vol. 177, No 2, pp. 305–319.

Chudik, A. and Pesaran, M. H. (2011), 'Infinite-dimensional VAR s and factor models', Journal of Econometrics, Vol. 163, No 1, pp. 4–22 .

Clements, M. P. and Hendry, D. F. (1999), 'Forecasting Non-stationary Economic Time Series', The MIT Press, London, England.

Dees, S., Di Mauro, F., Pesaran, M. H. and Smith, V (2007), 'Exploring the international linkages of the Euro Area: a global VAR analysis', Journal of Applied Econometrics, Vol. 22, pp. 1–38.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing Predictive Accuracy, Journal of Business and Economic Statistics', Vol. 13, pp. 253–263.

Diron, M. (2008), 'Short-term forecasts of euro area real GDP, growth: an assessment of real-time performance based on vintage data', Journal of Forecasting, Vol. 27, No 5, pp. 371–390.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), 'The Generalized dynamic Factor Model: Identification and Estimation', Review of Economics and Statistics, Vol. 82, pp. 540–554.

Geweke, J. (1978), 'The temporal and Sectoral Aggregation of Seasonally Adjusted Time Series', Seasonal Analysis of Economic Time Series, ed Zellner, A., Washington, DC: US. Department of Commerce, Census Bureau.

Giacomini, R. and Granger, C.W.J. (2004), 'Aggregation of Space-Time Processes', Journal of Econometrics, Vol. 118, pp. 7–26.

Giacomini, R. and White, H. (2006), 'Tests of conditional predictive ability', Econometrica, Vol. 74, pp. 1545–1578.

Giannone, D and Reichlin, L (2009), 'Comments on "Forecasting economic and financial variables with global VARs" ', International Journal of Forecasting, Vol. 25, pp. 684–686.

Giannone, D., Reichlin, L. and Small, D. (2008), 'Nowcasting: The Real Time Informational Content of Macroeconomic Data Releases', Journal of Monetary Economics, Vol. 55, pp. 665–76.

Granger, C. W. J. (1987), 'Implications of Aggregation with Common Factors', Econometric Theory, Vol. 3, No 02, pp. 208–222.

Hendry D. F. and Hubrich, K. (2011), 'Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate', Journal of Business and Economic Statistics, Vol. 29, pp. 216–227.

Kuzin, V., Marcellino, M. and Schumacher, C. (2011), 'MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area', International Journal of Forecasting, Vol. 27, No 2, pp. 529–542.

Lütkepohl, H. (1984), 'Forecasting Contemporaneously Aggregated Vector ARMA Processes', Journal of Business and Economic Statistics, Vol. 2, No 3, pp. 201–14.

Landefeld, J.S., Seskin, E. P. and Fraumeni, B. M. (2008), 'Taking the Pulse of the Economy: Measuring GDP', Journal of Economic Perspectives, Vol. 22, No 2, pp. 193–216.

Lui, S. and Mitchell, J. (2013), 'Nowcasting Quarterly Euro Area GDP Growth using a Global VAR Model', The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis, eds F. di Mauro and M.H. Pesaran, Oxford University PressForthcoming.

Lui, S. and Mitchell, J. (2013), 'Producing Early Euro-Area GDP Estimates using a Global VAR Model', Working Paper, EUI.

Marcellino, M. and Stock, J.H. and Watson, M.W. (2003), 'Macroeconomic Forecasting in the Euro Area: Country Specific versus Euro Area Information', European Economic Review, Vol. 47, pp. 1–18.

Mariano, R. S. and Murasawa Y. (2003), 'A new coincident index of business cycles based on monthly and quarterly series', Journal of Applied Econometrics, Vol. 18, No 4, pp. 427–443.

Mazzi, G-L. and Mitchell, J. and Montana, G. (2013), 'Density nowcasts and model combination: nowcasting Euro-area GDP growth over the 2008-9 recession', Oxford Bulletin of Economics and Statistics, Vol. 76, pp. 233–256.

Mitchell, J., Smith, R. J., Weale, R. M., Wright, S. and Salazar, E. L. (2005), 'An Indicator of Monthly GDP, and an Early Estimate of Quarterly GDP, Growth', Economic Journal, Vol. 115, No 501, pp. 108–129.

Pesaran, M. H. (2006), 'Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure', Econometrica, Vol. 74, No 4, pp. 967–1012.

Pesaran, M H. and Timmermann, A. (2007), 'Selection Of Estimation Window In The Presence Of Breaks', Journal of Econometrics, Vol. 137, No 1, pp. 134–161.

Pesaran, M. H. and Schuermann, T. and Smith, V (2009), 'Forecasting economic and financial variables with global VAR's', International Journal of Forecasting, Vol. 25, pp. 624–675.

Pesaran, M. H. and Schuermann, T. and Weiner, S. M. (2004), 'Modelling regional interdependence using a global error-correcting macroeconometric model', Journal of Business and Economic Statistics, Vol. 22, No 2, pp. 129–162.

Salazar, E. and Weale, M. (1999), 'Monthly data and short-term forecasting: an assessment of monthly data in a VAR, model', Journal of Forecasting, Vol. 18, pp. 447–462.

Stock, J.H. and Watson, M.W. (2002), 'Forecasting using principal components from a large number of predictors', Journal of the American Statistical Association, No 460, Vol. 97, pp. 1167–1179.

Stock, J H. and Watson, M W. (2002), 'Macroeconomic forecasting using diffusion indexes', Journal of Business and Economic Statistics, No 2, Vol. 20, pp. 147–162.

Taylor, J. B. (1978), 'Comments on " The temporal and Sectoral Aggregation of Seasonally Adjusted Time Series" ' By John Geweke, Seasonal Analysis of Economic Time Series, ed Zellner, A., Washington, D.C.: US. Department of Commerce, Census Bureau.

VII Big Data and Rapid Estimates

16 The Challenge of Big Data

Authors

Dario Buono
Eurostat
dario.buono@ec.europa.eu

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

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

Fotis Papailias
King's College, University of London and Quantf Research
fotis.papailias@kcl.ac.uk

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

16.1 Introduction

The recent economic crisis has emphasized the importance for policy-makers and economic agents of a real-time assessment of the current state of the economy and its expected developments, when a large but incomplete information set is available. The main obstacle is the delay with which key macroeconomic indicators such as GDP and its components, but also fiscal variables, regional/sectoral indicators and disaggregated data are released. For example, GDP data are only available on a quarterly basis and the first/flash estimate is only published with a 4-6 week delay or longer, depending on the country. Moreover, preliminary data are often revised afterwards, in particular around turning points of the business cycle.

On the other hand, a large and growing number of more timely leading and coincident indicators is available, at a monthly, daily or even higher frequency, based in particular on financial, survey and internet data, though sometimes subject to short samples, missing observations and other data irregularities. This has stimulated a vast amount of statistical and econometric research on how to exploit the large, timely and higher frequency but irregular information, usually referred as big data, to provide estimates for key low frequency economic indicators. A parallel, more empirical, literature has instead focused specifically on the use of big data for nowcasting economic indicators, often using rather simple econometric techniques and specific big data sources, mainly Google Trends.

Finally, a more theoretical literature has developed new, or adapted old, statistical and econometric methods to handle very large sets of explanatory variables, such as those associated with big data. Broadly speaking, they are based either on summarizing the many variables, or on selecting them, or on combining many small models, after a proper data pre-treatment. Some techniques are borrowed from machine learning, where prediction is of key interest but data are typically assumed to be i.i.d. rather than serially correlated and possibly with changing means and variances over time, so that these techniques have to be properly extended prior to use on economic data.

As a matter of fact, the potential of big data is largely unexplored even if there is a growing interest for their utilisation either in the regular production process of official statistics or as a complement of traditional information to construct rapid estimates. Based on the definitions adopted in chapter 2, we argue that big data has potential in the construction of flash estimates, however the current availability of big data is restricted and nowcasting examples limited. By contrast, the big data can be used to construct nowcasting or even advanced estimates as well as to provide regularly updated estimates such as at monthly and weekly frequencies of the target variables. For this reason, in this section we are mostly focusing on nowcasting and short-term forecasting exercises.

This chapter is organised as follows: Section 16.2 describes the different types of big data, section 16.3 analyses some advantages and drawbacks of big data while section 16.4 briefly reviews the existing big data literature. Section 16.5 shortly presents the modelling strategies for big data, section 16.6 contains some examples showing, even if in a very preliminary way, the potential of big data in nowcasting and short-term forecasting, while section 16.7 concludes.

16.2 Big Data types

The literature provides various definitions of big data. Part of the problem is that the meaning of big data can differ significantly across disciplines. One possibility to obtain a general classification is to adopt the "4Vs" classification, originated in IBM, which relates to:

- Volume (Scale of data),
- Velocity (Analysis of streaming data),
- Variety (Different forms of data) and
- Veracity (Uncertainty of data).

However, this classification seems too general to guide empirical nowcasting applications.

A second option is to focus on numerical data only, which can either be the original big data, or the result of a transformation of unstructured data, and refer to the size of the dataset. Unstructured data such as, e.g., credit card transactions or other data that describe the disaggregated actions or characteristics of many agents are likely to need to be transformed to a two dimensional panel structure where time is usually one dimension. Following, e.g. Doornik and Hendry (2015), we can distinguish three main types of big data:

- Tall
- Fat
- Huge

16.2.1 Big Data types: "Tall"

"Tall" means not so many variables, N, but many observations, T, with $T >> N$. This is for example the case with tick by tick data on selected financial transactions or search queries. In this case T is indeed very large in the original time scale, say seconds, but it should be considered whether it is also large enough in the time scale of the target macroeconomic variable of the nowcasting exercise, say quarters or months.

Tall datasets at very high frequency are not easily obtained, as they are generally owned by private companies. Moreover, they generally require substantial pre-treatment, as indicators typically present particular temporal structures (related, e.g., to market micro-structure) and other types of irregularities, such as outliers, jumps and missing observations.

16.2.2 Big Data types: "Fat"

"Fat" means many variables, but not so many observations, $N >> T$. Large cross-sectional databases fall into this category, which is not so interesting from an economic nowcasting point of view, unless either T is also large enough or the variables are homogeneous enough to allow proper model estimation (e.g., by means of panel methods) and nowcast evaluation. However, Fat datasets can be of interest in many other applications of big data, both inside official statistics, e.g., for surveys construction, and outside, e.g., in marketing or medical studies. Also, as the collection of big data started only rather recently, Fat datasets are perhaps the most commonly available type.

Actually, statistical methods for big data are mainly meant for Fat datasets, e.g., those developed in the machine learning literature, as they only require a large cross-section of i.i.d. variables. When a (limited) temporal dimension is also present, panel estimation methods are typically adopted in the economic literature, but factor based methods can also be applied. Classical estimation methods are not so suited, as their finite (T) sample properties are generally hardly known, while Bayesian estimation seems more promising, as it can easily handle a fixed T sample size and, with proper priors, also a large cross-sectional dimension.

16.2.3 Big Data types: "Huge"

"Huge" means many variables and many observations, i.e., very large N and T. This is perhaps the most interesting type of data in a nowcasting context even if, unfortunately, it is not so often available. Big data collection started only recently, while collection of the target economic indicators started long ago, generally as far back as the 1950s or 1960s for many developed countries. Google Trends, publicly available summaries of a huge number of specific search queries in Google, are perhaps the best example in this category, and not by chance the most commonly used indicators in economic nowcasting exercises.

Contrary to basic econometrics and statistics, in Huge datasets both T and N diverge and proper techniques must be adopted to take this feature into account at the level of model specification, estimation and evaluation. For example, in principle it is still possible to consider information criteria such as BIC or AIC for model

specification (indicator selection for the target variable in the nowcasting equation), although it is the case that modifications may be needed to account for the fact that N is comparable or larger than T, as opposed to much smaller as assumed in the derivations of information criteria.

Further, in the case of a linear regression model with N regressors, 2^N alternative models should be compared, which is not computationally feasible when N is very large, so that efficient algorithms that only search specific subsets of the 2^N possible models have been developed. Moreover, the standard properties of the OLS estimator in the regression model are derived assuming that N is fixed and (much) smaller than T. Some properties are preserved, under certain conditions, also when N diverges but empirically the OLS estimator does not perform well due to collinearity problems that require a proper regularization of the second moment matrix of the regressors.

This is in turn relevant for nowcasting, as the parameter estimators are used to construct the nowcast (or forecast). As a result of these problems for OLS, a number of regularisation or penalisation methods have been suggested. An early approach, referred to as Ridge regression, uses shrinkage to ensure a well behaved regressor sample second moment matrix. More recently, other penalisation methods have been developed. A prominent example is LASSO where a penalty is added to the OLS objective function in the form of the sum of the absolute values of the coefficients. Many related penalisation methods have since been proposed and analysed.

As an alternative to variable selection, the indicators could be summarized by means of principal components (or estimated factors) or related methods such as dynamic principal components or partial least squares. However, standard principal component analysis is also problematic when N gets very large, but fixes are available, such as the use of sparse principal component analysis. Finally, rather than selecting or summarizing the indicators, they could be all inserted in the nowcasting regression but imposing tight priors on their associated coefficients, which leads to specific Bayesian estimators.

16.3 Pros and Cons of Big Data for Macroeconomic Nowcasting

In a nowcasting context, we think of big data as complements rather than substitutes for more common coincident and leading indicators. Some of the problems discussed for internet based big data also apply to large datasets of conventional indicators. For example, collecting disaggregated macroeconomic and financial variables for an EU country easily leads to a few hundred indicators, and multiplying that for all the EU countries, e.g. in order to conduct a comparative analysis, leads to thousands of variables to be considered in formal econometric models, which can be hardly done with standard techniques.

Below we briefly discuss the different issues related to big data in the first subsection, while the advantages of big data are presented in the second subsection. Those aspects will be analysed in further detail in chapter 17 where a step-by-step approach for the construction of big data based nowcasting will be presented together with some recommendations.

16.3.1 Big Data issues

The first issue is the data availability. Most data pass through private providers and are related to personal aspects. Hence, continuity of data provision could not be guaranteed. For example, Google could stop providing Google Trends, or at least no longer make them available for free. Or online retail stores could forbid access to their websites to crawlers for automatic price collection. Or individuals could extend the use of software that prevents tracking their internet activities, or tracking could be more tightly regulated by law for privacy reasons.

The second one is the digital divide, or the fact that a sizable fraction of the population still has no or limited internet access. This implies that the available data are subject to a sample selection bias, and this can matter

for their use. Suppose, for example, that we want to nowcast unemployment at a disaggregate level, either by age or by regions. Internet data relative to older people or people resident in poorer regions could lead to underestimation of their unemployment level, as they have relatively little access to internet based search tools.

A third issue is that both the size and the quality of internet data keeps changing over time, in general much faster than for standard data collection. For example, applications such as Twitter or WhatsApp were not available just a few years ago, and the number of their users increased exponentially, in particular in the first period after their introduction. Similarly, other applications can be gradually dismissed or used for different uses. For example, the fraction of goods sold by EBay through proper auctions is progressively declining over time, being replaced by other price formation mechanisms.

Fourth, there is the bias in answers. Again more relevant for digital than standard data collection, is that individuals or businesses could not report truthfully their experiences, assessments and opinions. For example, some newspapers and other sites conduct online surveys about the feelings of their readers (happy, tired, angry, etc.) and one could think of using them, for example, to predict election outcomes, as a large fraction of happy people should be good for the ruling political party. But, if respondents are biased, the prediction could be also biased, and a large fraction of non-respondents could lead to substantial uncertainty.

A fifth issue is that data could not be available in a numerical format, or not in a directly usable numerical format. A similar issue emerges with standard surveys, for example on economic conditions, where discrete answers from a large number of respondents have to be somewhat summarized and transformed into a continuous index. However, the problem is more common and relevant with internet data. A final issue, again common also with standard data but more pervasive in internet data due to their high sampling frequency and broad collection set, relates to data irregularities like outliers, working days effects, missing observations and presence of seasonal/periodic patterns, etc., all of which require properly de-noising and smoothing the data.

16.3.2 Big Data advantages

Big data provide potentially relevant complementary information with respect to standard data, being based on rather different information sets. Moreover, they are timely available and, generally, they are not subject to subsequent revisions, all relevant features for potential coincident and leading indicators of economic activity.

Finally, they could be helpful to provide a more granular perspective on the indicator of interest, both in the temporal and in the cross-sectional dimensions. In the temporal dimension, they can be used to update nowcasts at a given frequency, such as weekly or even daily, so that the policy and decision makers can promptly update their actions according to the new and more precise estimates. In the cross-sectional dimension, big data could provide relevant information on units, such as regions or sectors, not fully covered by traditional coincident and leading indicators.

16.4 Literature Review

In this section, we summarise an extensive review of the most important research papers on big data, carried out by the authors in related work, in three distinct areas:

- Big Data in macroeconomics,
- Variable selection and dimensional reduction for big data in macroeconomics,
- Nowcasting in macroeconomics.

The aim of the review is to try to answer the following questions:

1. What are possible big data sources in relation to specific macroeconomic indicators (i.e. GDP, inflation and producer prices, employment and unemployment, industrial production index and retail trade deflated turnover)?

2. What are the advantages and disadvantages of the previously analysed sources?
3. What are the main types of statistical methods used in the big data in macroeconomics literature?
4. What are the possible gains generated either by the use of big data or new statistical methods or both in comparison with existing practices in the field of nowcasting?

Based on the surveyed papers, we can say that the use of Google Trends has been dominant in studies using big data in macroeconomics. There exist some papers based on Twitter data, also reviewed, but they are mainly in finance. Webscrapping and collection of online prices also offer some potential, especially for nowcasting inflation. However, such datasets are very difficult to obtain (and possibly sustain), even more so when many countries and long enough samples are required. A similar comment applies for credit card and financial transactions data, and for data summaries resulting from textual analysis.

From the literature it also emerges that the advantages of using data like Google Trends are:

- more timely forecasts, not subject to data revision,
- some improvements in forecast accuracy, even though these typically emerge with respect to simple benchmarks (AR models),
- ease of data access and collection,
- ease of data management and treatment,
- expected good data quality,
- reasonable likelihood that similar data will be available on a continuous basis and without major definitional changes.

There are also some disadvantages when using this data source. The first one being a typical sole use of such data, which can lead to biased results ("big data hubris"). Then there is the impossibility to access the raw data, and the lack of knowledge of the precise algorithms used to pre-treat and summarize them. Finally, the possibility exists that free access will be discontinued by the (private) data provider, or limited due to the introduction of more stringent privacy laws.

Varian (2014) provides an intuitive introduction in big data management and manipulation. Big data, possibly after transformation to some kind of numerical format, has to be stored in some sort of database, as it is difficult to be dealt with spreadsheets. A medium sized dataset (i.e. about a million observations) could be stored and manipulated using a relational database, such as MySQL, whereas a dataset of several million observations could be efficiently stored and manipulated by NoSQL databases. Sometimes, and depending on the nature of the research, a carefully selected subsample or summary of the data might be sufficient for analysis.

Hence, big data creation, storage and management typically require specific IT skills, software and hardware. The associated costs should be kept into consideration when assessing the potential benefits of big data for nowcasting.

In terms of statistics and econometrics, data analysis is typically broken down into four categories:

- pre-treatment and summarisation,
- estimation,
- hypothesis testing and
- prediction.

Since a large amount of data is available, penalised regressions such as LASSO, LARS, and elastic nets can be used instead of the standard linear or logistic regression. These techniques could also be used for

variable selection. Then, the choice of the final model should come from forecasting cross-validation so that the researcher makes sure the model has good out-of-sample predictive ability.

16.5 Big Data modelling

Let $y_t, t = 1, \ldots, T$, be the target variable and $x_t = (x_{1t}, \ldots, x_{Nt})'$ be a set of potential predictors, with N very large. We do not assume a particular data generating process for y_t but simply posit the existence of a representation of the form

$$y_t = a + g(x_{1t}, \ldots, x_{Nt}) + u_t, \tag{16.1}$$

which implies that $E(u_t|x_{1t}, \ldots, x_{Nt}) = 0$. While the potential nonlinearity in (1) might, in principle, be worth exploring, it is extremely difficult to model nonlinearities in the context of large datasets and no work is available on this in the big data literature. As a result, we consider an approximating linear representation of the form,

$$y_t = a + \sum_{i=1}^{N} \beta_i x_{it} + u_t \tag{16.2}$$

with u_t denoting a martingale difference process and where the set of $x_{it}s$ can also contain products of the original indicators in order to provide a better approximation to (1). Our main aim is to provide estimates for current and future values of y_t , where either no, or only a preliminary, value for y_t is available from official statistics. To do so, we can rely on many approaches, which can be categorised in three main strands.

The first strand aims to provide estimates for $\beta = (\beta_1, \ldots, \beta_N)'$. While ordinary least squares (OLS) is the benchmark method for doing so, it is clear that if N is large this is not optimal or even feasible (when $N > T$).

So other methods need to be used. We consider two classes of methods: the first one is sparse regression, with origins in the machine learning literature. A main aim there is to stabilise the variability of the estimated β_i. The second class considers the use of a variety of information criteria such as AIC or BIC to select a smaller subset of all the available predictors. The second strand consists of reducing the dimension of x_t by producing a much smaller set of generated regressors, which can then be used to produce nowcasts and forecasts in standard ways. The third strand suggests the use of a (possibly very large) set of small models, one for each available indicator or small subset of them, and then the combination of the resulting many nowcasts or forecasts.

Our main findings indicate that Multiple Testing (related to Boosting) and, potentially, some variant of LASSO could be the preferred approaches among the set of machine learning techniques. Among the data reduction techniques, PCA and, possibly PLS are promising (possibly in their "sparse" versions). It is also worthwhile experimenting with Bayesian regression, with substantial shrinkage (specific priors). Forecast combination, with simple equal weighting, could be also useful, though not so explored in big data context.

Finally, all these approaches should be also modified to take into account the possibility of a different timing for the target and the indicator variables. We have surveyed a number of alternative methods to handle mixed frequencies, and it turns out that Unrestricted MIDAS or bridge modelling appear as the most promising approaches, as they preserve linearity and do not add an additional layer of computational complexity.

16.6 Empirical Analysis

We consider nowcasting and short-term (one- to twelve- months ahead) forecasting three key macroeconomic variables:

1. Inflation (measured by the growth rate in the Consumer Price Index (CPI)),
2. Growth in retail sales (measured by the growth rate of the Retail Trade Index (RTI)), and
3. Unemployment Rate.

The exercise is conducted recursively in a pseudo out of sample framework, using monthly data for three economies: Germany (DE), Italy (IT) and the UK. We assess the relative performance of Big Data (proxied via weekly Google Trends) and standard indicators (based on a large set of weekly and monthly economic and financial variables). In fact, as we have mentioned several times, we think of Big Data as providing complementary information, and we wish to assess how useful it is in a forecasting context relative to standard indicators. We also evaluate the role of several econometric methods and alternative specifications for each of them (with or without big data), also capable of handling the frequency mismatch in our data. Specifically, we consider:

- Naive and autoregressive (AR) models as benchmarks
- Dynamic Factor Analysis (DFA)
- Partial Least Squares (PLS)
- Bayesian Regression (BR) and
- LASSO regression

DFA and PLS are representatives of data reduction methods; BR and LASSO are representatives of, respectively, econometric and machine learning techniques. In addition, we also consider model averaging. Overall, we have a total of 255 models and model combinations.

Further, we separate our data in two groups. The first set of data is made up of standard macroeconomic and financial indicators (MacroFin) which the second is obtain from Google Trends (Google). We try to see how Google adds value to the standard MacroFin datasets as predictors.

A full description of our results is beyond the scope of this summary chapter. However, it is of great interest to compare the performance of the best models obtained from this work to official Eurostat flash estimates. In order to make the comparison as accurate as possible we proceed as follows.

1. The error is calculated using levels for RTI and Unemployment and the one period growth (MoM) for the CPI. The latter is necessary in order to make the results comparable to Eurostat's estimates which are offered in MoM.
2. The evaluation period for all models (including Eurostat's estimates) is set to 2013-Jan to 2016-Jan for all series apart from UK Retail Trade, which is 2013-Aug to 2016-Jan, and UK Unemployment Rate, which is 2013-Jan to 2015-Dec. This also allows for cross-comparison of results in this subsection and in the previous one.
3. As we do not account for data revisions, we use the time series of the initial ("flash") estimate of Eurostat for each reference month ignoring further revisions.
4. In an effort to shed extra light on the usefulness of big data for nowcasting, we report the differences of the MAE of the best $MacroFin + Google$ and $MacroFin$ models to the MAE of the official nowcasts (Table 16.1).[1]

[1] We use the difference instead of the fraction because in some cases Eurostat's estimates carry no error.

In addition to the above tables we also include Figure 16.1 where we plot the time series of the difference of the absolute error of the best $MacroFin + Google$ and the absolute error of Eurostat nowcasts.

Starting with the CPI MoM series, from the top panel of Table 16.1 we see that for IT the difference of $MacroFin + Google$ and $MacroFin$ models to the official nowcasts is very close to zero, indicating that the models perform similarly. The positive values for UK and DE indicate that the official nowcasts have smallest MAE for these countries. For DE we see a slight improvement of the model which includes Google Trends (0.293 compared to 0.302), however for the UK we see no difference.

Moving to the RTI series, from the mid panel of Table 16.1 we see that for DE the difference is negative, indicating that the suggested models outperform the official nowcasts. For IT, the best model which includes Google Trends improves the performance, which is still worse than that of the official flash estimate but very close to it. For UK we again see no difference from the inclusion of the Google trends, and a clear dominance of the official estimates.

For the unemployment rate, in the case of Germany and Italy we see that our best models provide a smaller MAE compared to the official nowcasts, as the MAE difference is negative, though still close to zero. For the UK again we have no improvement.

Interestingly, from Figure 16.1 it emerges that even when the official flash estimates are on average more accurate than those based on our models there are periods where the ranking is reversed, suggesting that a combination of the two approaches could perform even better.

Finally, in Table 16.2 we regress the official Eurostat nowcast errors on the forecasts of our best models. The idea is that if we can explain the official errors using our nowcasts, than the latter could be also used to improve the official flash estimates. As expected, based on the previous analysis, the regressions confirm the result for the Italian Unemployment Rate, indicating significance at 10% level, while the other coefficients are not statistically significant, in line with the small differences in MAE reported above.

In summary, this analysis shows that our nowcasts (also those based on Google trends) are generally not better than the official flash estimates, but in some cases like the unemployment rate of DE and IT, perform better on average. It should be stressed that the alternative estimates are based on different information sets and timing, so that they are not really comparable. Moreover, the comparison is based on a very short sample, with few variables and countries (and a specific set of summary big data indicators). Hence, the main conclusion of this subsection is that a more detailed empirical comparison and evaluation of the Eurostat and big data based nowcasts, while challenging and time demanding, would be quite interesting.

Table 16.1: Comparing the best methods to Eurostat. Reporting differences of nowcasting MAE. A *negative* value indicates that the nowcast error of the underlying model is smaller than Eurostat's. The results use the smaller evaluation sample which corresponds to the Eurostat official nowcasts. The evaluation periods are: 2013-Aug to 2016-Jan (30 months) for the UK, Retail Trade, 2013-Jan to 2015-Dec (36 months) for UK, Unemployment rate and 2013-Jan to 2016-Jan (37 months) for all other series.

Consumer Price Index

	DE	IT	UK
MacroFin+Google	0.293	0.003	0.267
MacroFin	0.302	0.004	0.267

Retail Trade Index

	DE	IT	UK
MacroFin+Google	-0.017	0.078	0.480
MacroFin	-0.015	0.101	0.484

Unemployment Rate

	DE	IT	UK
MacroFin+Google	-0.020	-0.059	0.134
MacroFin	-0.018	-0.058	0.151

Table 16.2: Regressing the error of the Eurostat nowcasts on the best $MacroFin + Google$ and $MacroFin$ only model. The evaluation periods are: 2013-Aug to 2016-Jan (30 months) for the UK, Retail Trade, 2013-Jan to 2015-Dec (36 months) for UK, Unemployment rate and 2013-Jan to 2016-Jan (37 months) for all other series. $*$ indicates significance at 10% level.

Consumer Price Index

	DE	IT	UK	DE	IT	UK
	MacroFin+Google			**MacroFin**		
$\hat{\alpha}$	0.001	-0.057	NA	0.005	-0.059	NA
$\hat{\beta}$	0.033	0.298	NA	0.004	0.304	NA

Retail Trade Index

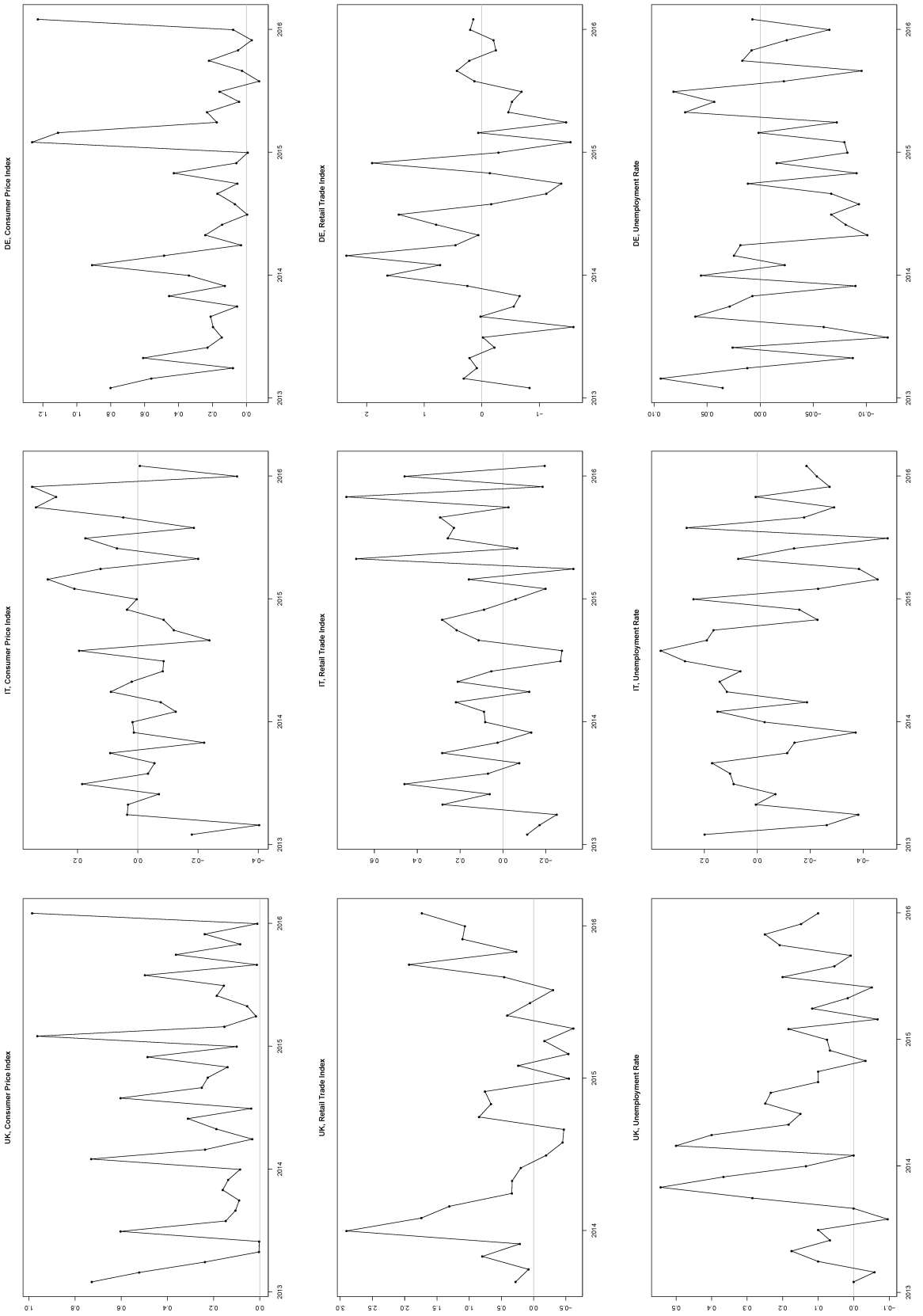
	DE	IT	UK	DE	IT	UK
	MacroFin+Google			**MacroFin**		
$\hat{\alpha}$	-6.429	8.677	0.363	-6.256	10.317	0.384
$\hat{\beta}$	0.062	-0.092	-0.004	0.060	-0.110	-0.004

Uneployment Rate

	DE	IT	UK	DE	IT	UK
	MacroFin+Google			**MacroFin**		
$\hat{\alpha}$	0.277	2.733*	-0.14047	0.273	2.864*	-0.14641
$\hat{\beta}$	-0.060	-0.233*	0.02121	-0.05921	-0.243*	0.02204

$MacroFin + Google$: The best model (MAE) using macro/financial variables and Google Trends.
$MacroFin$: The best model (MAE) using macro/financial variables.

Figure 16.1: Reporting the difference of absolute nowcast error of best MacroFin+Google model and Eurostat. Positive numbers indicate that the error of the Macro+Google model is larger. Negative values indicate that the error of Eurostat nowcast is larger.



16.7 Conclusions

Overall, our suggestion is to take a pragmatic approach that balances potential gains and costs from the use of Big Data for nowcasting macroeconomic indicators, in addition to standard indicators. A preliminary step should be an a priori assessment of the potential usefulness of Big Data for a specific indicator of interest, such as GDP growth, inflation or unemployment. This requires evaluating the quality of the existing nowcasts and whether any identified problems, such as bias or inefficiency or large errors in specific periods, can be fixed by adding information as potentially available in Big Data based indicators. Similarly, it should be considered whether these additional indicators could improve the timeliness, frequency of release and extent of revision of the nowcasts. Relevant information can be gathered by looking at existing empirical studies focusing on similar variables or countries.

Bibliography

Doornik, J. A., Hendry, D. F. (2015). "Statistical Model Selection with Big Data", *Cogent Economics & Finance*, 3(1).

Varian, H. R. (2014). "Big Data: New Tricks for Econometrics", *Journal of Economic Perspectives*, 28(2), 3-28.

17

Guidance and Recommendations on the Use of Big data for Macroeconomic Nowcasting

Authors

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

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

Fotis Papailias
King's College, University of London and Quantf Research
fotis.papailias@kcl.ac.uk

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

17.1 Introduction

Parallel advances in IT and the social use of Internet related applications make nowadays a vast amount of information available in real time. The associated Big Data are potentially very useful for a variety of applications, ranging from marketing to tapering fiscal evasion.

From the point of view of statistical institutes, the main question is whether and to what extent Big Data are valuable to expand, check and improve the data production process. In this paper we focus on the particular case of using Big Data for macroeconomic nowcasting, thus potentially enhancing the timely availability and precision of early estimates of key macroeconomic variables, and possibly providing new Big Data based coincident and leading indicators of economic activity. While clearly potentially of important value in this context, Big Data raise a number of old and new issues, related to data access, preparation, cleaning, filtering, and evaluation, which are all the more relevant in the context of official statistics.

Big Data also require proper statistical and econometric modelling and evaluation techniques. A range of methods are available, but they should be used with care as they have been developed either (in the econometric literature) to handle large but not really big datasets of time series, or (in the statistical and machine learning literatures) to model really big datasets of i.i.d observations, while we would like to work with big sets of time series. Broadly speaking, the available methods are based either on summarizing the many variables, or on selecting them, or on combining many small models, after a proper data pre-treatment. Some indications on the most promising approaches can emerge from a careful and critical review of the existing empirical literature on the use of big data for nowcasting economic indicators, which is however still in its infancy, or from pilot empirical studies.

Given this background, in this chapter we provide a step by step approach, accompanied by specific recommendations for the use of big data for macroeconomic nowcasting. This step-by-step approach is intended to guide the researcher from the identification and the choice of Big Data to its pre-treatment and econometric modelling until the comparative evaluation of results to obtain a very useful tool for decision about the use or not of Big Data.

More specifically, Step 1 considers Big Data usefulness within a nowcasting exercise. Step 2 discusses issues related to the design of the exercise: possible big data sources, definition of the type of variables/information to be extracted from big-data, etc. Step 3 assesses big-data accessibility and the quality of envisaged big-data sources. Step 4 focuses on Big data preparation and Step 5 on the design of the econometric modelling strategy. Step 6 deals with the evaluation of the results (choice of benchmark model or nowcasts, evaluation metrics, criteria for assessing the significance of the accuracy and/or timeliness gains, etc.). Step 7 considers the overall implementation of big data based nowcasting. All these steps are discussed in Section 17.2, and for each of them we provide specific recommendations. Section 17.3 provides a more compact discussion and concludes.

17.2 A step-by-step approach

17.2.1 Step 1: Big Data usefulness within a nowcasting exercise

Description

The advantages of Big Data have been widely emphasized both in the general press and in more specialized journals and meetings. In a nowcasting context, Big Data provide potentially relevant *complementary* information with respect to standard data, being based on rather different information sets. Moreover, it is timely available and, generally, not subject to subsequent revisions, all relevant features for potential coincident and leading indicators of economic activity. Further, it can provide a more granular perspective on the indicator of interest, both in the temporal and in the cross-sectional dimensions. In the temporal dimension, Big Data can be used to update nowcasts at a given frequency, such as weekly or even daily, so that the policy and

decision makers can promptly update their actions according to the new and more precise estimates. In the cross-sectional dimension, Big Data could provide relevant information on units, such as regions or sectors, not fully covered by traditional coincident and leading indicators. All these benefits should be however contrasted with the possible drawbacks, which are many but all controllable. In fact, clearly, the collection and preparation of Big Data based indicators is far more complex than that of standard coincident and leading indicators, which are often directly downloadable in ready to use format from the web through statistical agencies or data providers. Following the guidelines provided in this section should permit to form a reasonable expectation of the costs of Big Data and of whether it will lead to additional gains in terms of nowcasting performance. Obviously, the relevance and the usefulness of Big Data can vary across the indicators so that they have to be checked in a case by case way. Finally, it is important to notice as the potential of Big Data has to be evaluated either in terms of the ability of improving the precision of nowcasting at a given time lag or in terms of the possibility of producing timelier nowcasts without significant deterioration of their reliability.

Recommendations

- Suggest the use of Big Data only when there are well founded expectations of their usefulness either for fixing problems in existing nowcasting or to improve the timeliness and/or the quality of nowcasting.
- Do not consider Big Data sources with doubtful or even spurious correlations with the target variable.

17.2.2 Step 2: Big Data search

Description

The starting point for an assessment of the potential benefits and costs of the use of Big Data for macroeconomic nowcasting is the identification of their source. A first main provider is represented by **Social Networks** (human-sourced information), broadly defined to include proper social networks such as Facebook, Twitter, Tumblr etc., but also blogs and comments, pictures (Instagram, Flickr, Picasa etc.), videos (Youtube, etc.), Internet searches, mobile data content (text messages, GPS positioning, user generated maps, etc.), e-mails. The associated data is, typically, loosely structured and often ungoverned.

A second main source of Big Data are **Traditional Business Systems** (process-mediated data). These processes record and monitor business events of interest, such as registering a customer, manufacturing a product, taking an order, etc. The process-mediated data thus collected by either private or public institutions is highly structured and includes transactions, reference tables and relationships, as well as the metadata that sets its context. Traditional business data is the vast majority of what IT managed and processed, in both operational and Business Intelligence systems. Usually structured and stored in relational database systems, including also “Administrative data", it can be further grouped into data produced by Public Agencies (medical records, social insurance, etc.) and data produced by businesses (commercial transactions, banking/stock records, e-commerce, credit cards, etc.).

A third, fast expanding, provider of Big Data is the so-called **Internet of Things** (machine-generated data). This data is derived from sensors and machines used to measure and record the events and situations in the physical world. Examples include data from fixed sensors (home automation, weather/ pollution sensors, traffic sensors/ webcam, etc.) or mobile sensors (tracking: mobile phone location, cars, satellite images, etc.), or data from computer systems (logs, web logs, etc.). The well-structured nature of machine-generated data makes it suitable for computer processing, but its size and speed is way beyond traditional approaches.

From an economic nowcasting point of view, all the three types of big data are potentially relevant. For example, selected internet searches and/or “tweets"[1] (Social Networks), credit card transactions (Traditional Business Systems), or number of navigating commercial vessels in a certain area (Internet of Things) could all provide useful leading indicators for GDP growth of a country. Hence, a crucial step for a proper use of Big Data for nowcasting is a *careful search and classification of existing data*. A priori, it is very difficult to give

[1] Twitter-generated messages.

general guidelines on a preferred data source because the choice is heavily dependent on the target indicator of the nowcasting exercise. Big Data users should be aware that the type of Big Data selected can heavily influence the whole nowcasting exercise both in terms of Big Data handling and Big Data modelling due to their more or less unstructured typology and to their size.

Recommendations

- Searching in the wider possible set of Big Data having clearly in mind the specificities and the characteristics of the target variable as well as what we want to nowcast.
- Checking for the adherence of available Big Data to what the target variable is really measuring.

17.2.3 Step 3: Assessment of big-data accessibility and quality

Description

Unfortunately, not all existing Big Data are also available. In fact, most data pass through private providers and are related to personal aspects. Hence, once a potentially useful Big Data source is damnified for nowcasting a specific indicator of interest, it should be evaluated *whether and at what cost the information is actually available*.

Besides availability, an additional issue is continuity of data provision, which could not be guaranteed. For example, Google could stop providing Google Trends, or at least no longer make them available for free. Or online retail stores could forbid access to their websites to crawlers for automatic price collection. Or individuals could extend the use of softwares that prevent tracking their internet activities, or tracking could be more tightly regulated by law for privacy reasons. Continuity of data availability is more an issue for the use of internet data in official statistics than for a pure nowcasting purpose, as it often happens in nowcasting that indicators become unavailable or no longer useful and must be replaced by alternative variables. That said, *continuity and reliability of provision are important elements* for the selection of a big data source.

A Big Data feature specifically relevant for nowcasting applications is the overall *number of temporal observations in the frequency of the target economic indicator* (typically, months/quarters). Unfortunately, this is generally low, even if in high frequency or cross-sectionally there can be thousands of observations, as Big Data generation and collection has started only recently. A short temporal sample is problematic as the Big Data based indicators need to be related to the target low frequency macroeconomic indicators and, without a long enough sample, the parameter estimators can be noisy and the ex-post evaluation sample for the nowcasting performance too short. On the other hand, several standard informative indicators, such as surveys and financial condition indices, are also only available over short samples, starting after 2000, and this feature does not prevent their use. Moreover, panel studies using rather short time series but for many countries could be also informative and alleviate the need for long temporal Big Data samples. Furthermore, the availability of a good quality documentation or of a regularly updated metadata associated to Big Data must be viewed as an important characteristic when analysing and selecting available Big Data.

In the context of quality evaluation of Big Data, for nowcasting exercises, it can be also relevant to conduct a careful *evaluation of the presence of a possible bias*. A first source of bias can be related to the “digital divide", the fact that a sizable fraction of the population still has no or limited internet access. For example, if we want to nowcast unemployment at a disaggregate level, either by age or by regions, using internet search data (e.g., Google Trends) could lead to underestimation of the unemployment level of older people or people resident in poorer regions, as they have relatively little access to internet based search tools. A second source of bias emerges when individuals or businesses do not report truthfully their experiences, assessments and opinions. For example, some newspapers and internet sites conduct online surveys about the feelings of their readers (happy, tired, angry, etc.) and one could think of using them, for example, to predict election outcomes, as a large fraction of happy people should be good for the ruling political party. But, if respondents are biased, the prediction could be also biased, and a large fraction of non-respondents could lead to substantial uncertainty.

A final issue specifically relevant for Big Data nowcasting is the *control of the stability of the relationship with the target variable*. This is a common problem also with standard indicators, as the type and size of economic shocks that hit the economy vary over time. In the case of Big Data an additional potential source of instability is the fact that both their size and quality keeps changing over time, in general much faster than for standard data collection. For example, applications such as Twitter or WhatsApp were not available just a few years ago, and the number of their users increased exponentially, in particular in the first period after their introduction. Similarly, other applications can be gradually dismissed or used for different uses. For example, the fraction of goods sold by EBay through proper auctions is progressively declining over time, being replaced by other price formation mechanisms. The presence of the bias and the instability of the relationships between the target variables and the Big Data can be seen as discriminants for the use of Big Data. If both can be appropriately modelled, then Big Data can be used while if modelling techniques fail in dealing with them, the use of the concerned Big Data should be discouraged.

Recommendations

- Privileging data providers which are able to give sufficient guarantee of the continuity of the data process and of the availability of a good and regularly updated metadata associated to the Big Data
- Privileging Big Data sources which ensure sufficient time and cross-sectional coverage to properly building up a nowcasting exercise.
- If a bias in the Big Data answers is observed, provided that it has been reasonably stable in the last few years, a bias correction can be included in the nowcasting strategy.
- If a bias in the Big Data answers is very unstable, then the Big Data should be considered not reliable enough to be used in a nowcasting exercise.
- In order to deal with a possible instability of the relationships between the Big Data and the target variables, nowcasting models should be re-specified on a regular basis (e.g. yearly) and occasionally in presence of unexpected events.

17.2.4 Step 4: Big data preparation

Description

Even if available, there are often substantial costs to make the Big Data suitable for nowcasting exercises. Actually, big data is often unstructured, so that a proper mapping into regularly spaced cross-sectional or time series observations is required. For example, consider the use of internet search queries for nowcasting a specific indicator. This requires to select a given set of queries related to that indicator (including the country coverage, language, etc.), to collect information on the number of queries for each specific keyword in a given time span (e.g., in a day, week or month), and to aggregate all the keyword specific queries into an index that, ideally, should lead the specific target indicator.

This topic is clearly relevant but not considered formally in the econometric or statistical literature on Big Data. We have offered original insights on how to formalize this transformation process of unstructured data into time series. Unfortunately, there is no unique way to transform unstructured into structured data, as the specific transformation depends on the type of Big Data. However, most transformations can be treated in a unified analytical context, where they are considered as functions that map the Big Data into the real space. We have also illustrated and evaluated the forecasting ability of our proposal using a set of simulation experiments, which support its use. As the method is similar to what implemented by Google in the construction of Google Trends, the latter are also promising big data summaries, at least from a forecasting point of view.

Moreover, from the simulation experiments it turned out that not only the Big Data based indicators but also autoregressive terms were always present in the best forecasting models, which supports our suggested

empirical strategy of combining big data with autoregressive terms and other more standard macroeconomic and financial indicators.

Even when already available in numerical format or after their transformation into numerical form as in the previous step, pre-treatment of the Big Data is often needed to remove deterministic patterns such as outliers and calendar effects and deal with data irregularities, like missing observations. Furthermore, seasonal and non-seasonal short-term movements (i.e. infra-monthly ones) should be removed accordingly to the characteristic of the target variable. Data irregularities can be treated in two main ways in econometrics. A first possibility is to extend the econometric models with proper variables that account for specific data features. For example, dummy variables can handle missing observations and outliers, and seasonal dummies can proxy for general seasonal patterns in the data. Testing for the significance of these additional variables, often conducted by robust versions of $t-$ and $F-$tests, provides a useful tool to formally evaluate the extent and significance of the data irregularities. The extended model approach works rather well in the context of standard linear regression models with few regressors. However, when the additional dummies are added either in nonlinear models (such as factor models) or in models with a very large number of regressors which are later estimated by shrinkage methods, the performance is unclear and mostly so far unexplored.

In these cases, the second approach to handle data irregularities is suggested. This requires to “clean" the variables prior to econometric modelling, replacing outliers and missing observations with reasonable estimates, removing other deterministic effects (such as calendar ones) and filtering for seasonal and other short-term periodic movements such as intra-monthly or intra-weekly ones. When the number of variables is really large and/or the adjustment has to be implemented many times, as in the case of recursive forecasting exercises, it is convenient to work on a series by series basis. Since not all seasonal and calendar adjustment methods can be applied when data are available at high frequency, appropriate adjustment techniques need to be identified when the data are available at high frequency. The size of the datasets suggests resorting to robust and computationally simple univariate approaches.

In summary, a careful *assessment of aspects related to Big Data transformation and preparation* is needed prior to a nowcasting exercise, as they can be very time and resource demanding.

Recommendations

- Creating a Big Data specific IT environment where the original data are collected and stored with associated routines to automatically convert them into structured, either cross-sectional or time-series datasets.
- Ensure the availability of an exhaustive documentation of the Big Data conversion process.
- Whenever possible, all the data treatment (outliers correction, trading day correction, seasonal filtering, etc.) described in this step should be done within a unique framework in order to avoid inconsistencies between different parts of the process.
- The filtering of Big Data should be consistent to the one used for the target variables: for example if the target variable is not seasonally adjusted, there is no reason to remove the seasonal component from Big Data and vice-versa.

17.2.5 Step 5: Designing a Big Data modelling strategy

Description

Once temporally structured, properly cleaned, Big Data is available, we can proceed to identify and implement one or several proper econometric methods to match the target indicator with the Big Data based explanatory variables, and conduct a careful in-sample and pseudo-out-of-sample evaluation (cross-validation) of the alternative methodologies. We discuss these aspects in the next two steps.

Big Data prevents the use of standard econometric methods. For example, when the number of regressors is larger than that of observations ($N \ggg T$, as in FAT datasets), OLS estimation clearly cannot be used, as well as OLS based statistics, such as t-tests and F-tests to check the significance of regressors. Moreover, selecting regressors by means of information criteria also becomes not doable, as 2^N models should be compared, a number larger than one million already for $N = 20$ regressors. Furthermore, standard statistical theory to prove econometric properties such as unbiased and consistency of the estimators typically relies on fixed N and T diverging asymptotics (suited for TALL datasets, where $T \ggg N$). Instead, with big (potentially HUGE) data both N and T diverging asymptotics is needed, which is much more complex. While Big Data is typically associated with internet, as we have seen, in practice we have similar econometric issues also when pooling together standard economic indicators for many countries or sectors or firms or households, or with tick by tick data on selected financial transactions, which easily produces datasets with thousands of variables and/or thousands of observations for each variable.

A common approach is to either aggregate the data or to impose strong a priori assumptions on the econometric models for the disaggregate data, e.g., in panel data models the effects of the regressors are typically assumed to be the same for all units, only intercepts can differ, and the units do not interact among themselves. Clearly, in general these assumptions are invalid, and data aggregation leads to a loss of information. Another alternative for Fat and Huge datasets is to split them into smaller ones, each with a limited N dimension, apply standard techniques to each sub-dataset, and then re-group the results. However, also in this case cross-datasets linkages can be lost, as well as efficiency. Hence, *proper Big Data econometrics is needed*.

Big data econometrics has received a big boost in the recent past, more in terms of estimation and testing than forecasting. There are many approaches available, which we have categorised in five main classes that we now briefly summarize, providing guidelines on when to use each of them.

Machine Learning methods

In, machine learning methods, which are particularly suited for FAT datasets, the starting point is to somewhat regularize OLS estimation to make it feasible also when N is very large. This is typically achieved by adding a set of (nonlinear) constraints on the model parameters, which are thus shrunk towards pre-specified values, preferably towards zero in order to achieve a more parsimonious specification.

This class includes methods such as Penalised Regression, Ridge Regression, LASSO Regression, Adaptive LASSO, Elastic Net, SICA, Hard Thresholding, Boosting and Multiple Testing. Unfortunately, few applications of these methods are available in the context of macroeconomic nowcasting and forecasting, and even fewer cross-method comparisons. However, considering that we need methods that are robust and can handle temporally correlated and not only iid variables, and need to reach a balance between flexibility and computational time, Multiple Testing emerges as a promising candidate within this class.

Heuristic Optimisation

The rationale of Heuristic Optimisation is to use information criteria to reach a good balance between model fit and parsimony by assigning a penalty dependent on the number of model parameters (which is equal to that of regressors in the linear context). However, as mentioned, direct minimization of the information criteria is not doable when N is large, as the number of models to be compared is too large. Yet, the search for the best model could be restricted to specific regions of the vast model space, and Heuristic Optimisation provides methods for doing so.

Within this class of algorithms, it is worth mentioning Simulated Annealing, Genetic Algorithms, and MC^3. As the methods are iterative, and sometimes simulation based, they can become computationally very demanding when N is really large, say already about 1000. As they should be applied recursively in a macroeconomic forecasting context, not only for forecast evaluation but also for cross-validation, the computational aspect can become quickly prohibitive. However, in other contexts, e.g. for the construction of coincident and leading composite indices, these methods could be quite useful at the stage of indicator selection.

Dimensionality reduction techniques

A third class of econometric methods to properly handle big data is based on the idea of reducing the dimension of the dataset by producing a much smaller set of generated regressors, which can then be used in a second step in standard econometric models to produce nowcasts and forecasts in common ways.

There are naturally many ways to carry out dimensionality reduction, the most common are Principal Component Analysis and Partial Least Squares, which can handle TALL datasets, and Sparse Principal Component Analysis, which is also suited for FAT and HUGE datasets. Principal components were typically considered as statistical objects of little economic interest. However, recent research has shown that they can be also used as estimators for the unobservable factors in approximate dynamic factor models, where each of the large set of variables is driven by few (temporally correlated) common factors plus an idiosyncratic component, and there can be limited cross-sectional correlation among the idiosyncratic components. As factor models are often implied by economic theory, e.g., in DSGE models there are a few main economic shocks driving all variables, principal components have now also a stronger economic justification. However, they still have the feature of summarizing the information in a large dataset without targeting it to explain a specific variable, which could enhance the forecasting performance. Partial least squares permits to do this, with some additional computational costs.

Shrinkage Estimators and Bayesian Methods

Shrinkage estimators typically regularize OLS estimation, making it feasible also when N is very large and larger than T, by adding a set of a priori constraints on the model parameters. The constraints induce a bias in the estimator but also reduce its variance. As a consequence, some bias is also introduced in the forecasts, but their efficiency can be improved.

In a classical context, shrinkage can be obtained in various ways, including by means of stochastic constraints on the model parameters. However, Bayesian methods are more commonly used, where prior distributions on the parameters permit to achieve the desired level of shrinkage (besides allowing the user to incorporate a priori opinions on the relevance of each explanatory variable).

Bayesian regression has indeed attracted much interest recently also among practitioners, mostly due to the advancements in computing power that make simulation methods substantially faster than twenty years ago. Still, for Bayesian regression to be doable in a big data context, we need analytical rather than simulation based estimators. This can be achieved with a proper choice of prior distributions for the parameters, which should be conjugate (in order to have analytically computable posteriors of the same type as the priors), and imposed with sufficient tightness to guarantee a substantial amount of shrinkage (as in the penalized regressions).

Nowcast pooling

Forecast pooling (or combination) has a long tradition of empirical success, and nowcast pooling is promising as well. Possible reasons for the good performance of forecast pooling may be model misspecification, model uncertainty and parameter non-constancy, which are attenuated by weighting. As these features are likely present when modelling with big data, forecast combination could be helpful also in this context.

Hence, an alternative procedure in the presence of a big set of potentially useful leading indicators for the target variable of interest is to use a (possibly very large) set of small econometric models to produce nowcasts, one model for each of the N available indicator or small subset of them, and then to combine the resulting many nowcasts or forecasts into a single prediction. An additional benefit of this approach is an evaluation of the forecasting ability of each of the many available indicators, which is of interest by itself.

Sophisticated techniques for forecast combination have been suggested recently, including Bayesian model averaging and information theoretic weighting. However, both methods can be hardly applied for computational reasons when the number of models (N in case of a model for each variable) grows large. Instead,

simple weighting schemes, and in particular equal weighting, can be implemented also with very large N, and anyway with smaller N they often perform equally or better than more sophisticated alternatives.

Finally, it is worth mentioning that all these approaches for nowcasting with Big Data should be also modified to take into account the possibility of a different timing for the target and the indicator variables. For example, GDP and its deflator are only available on a quarterly basis, while Google trends are released on a weekly basis, and other internet based indicators, such as credit card transactions, are in principle available in real time. A similar issue occurs with standard indicators, such as surveys or financial variables, and a number of alternative econometric methods to handle mixed frequencies have been developed. Among them, Unrestricted MIDAS or bridge modelling appear as the most promising approaches in the presence of Big Data, as they preserve linearity and do not add an additional layer of computational complex.

Recommendations

- In the absence of any a priori information on the relative performance of various techniques, as many methods as possible should be evaluated and compared in a nowcasting context in order to select the best performing one.
- Alternative modelling strategies should be compared also by looking at the balance between their complexity in computational terms and their empirical performance.
- In case of mixed frequency data, linear methods such as UMIDAS and, as a second best, Bridge, should be privileged.
- Forecast combination and model averaging techniques, also when the mixed frequency aspect is present, can be used as an alternative to a large-scale comparison among competing techniques.

17.2.6 Step 6: Results evaluation of Big Data based nowcasting

Description

As Harford (2014) put it: " 'Big data' has arrived, but big insights have not. The challenge now is to solve new problems and gain new answers – without making the same old statistical mistakes on a grander scale than ever." This is the goal of this step: assessing in a critical and comprehensive way the contribution of Big Data for nowcasting an indicator of interest. This can be done, whenever possible within a real-time or a quasi real-time simulation exercise, by answering a set of questions, that we discuss in the next subsections.

Is there a "Big Data hubris"?

"Big data hubris" is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis, see Lazer et al. (2014). In a nowcasting context, this is the case for studies that are purely based on Big Data indicators, for example trying to anticipate unemployment using Google Trends only. It is not surprising that Big Data indicators are useful when used this way, but their usefulness could be spurious.

To attenuate the Big Data hubris we should think, as mentioned, of Big Data based indicators as complements to existing soft and hard data-based indicators, include all of them in econometric models, and assess the marginal contribution of each type of indicators. This evaluation should be conducted in an out-of-sample context, as in-sample Big Data can lead to overfitting.

Various specifications including alternative combinations of traditional and Big Data based indicators should be used for nowcasting over a training sample and their associated forecast errors evaluated and compared, for example in terms of mean squared and mean absolute forecast errors.

Is there a risk of “False positives"?

This question can be reformulated as: can we get some Big Data based indicators that nowcast well just due to data snooping? Similarly, can we get positive results because of model snooping, since we have seen that various econometric approaches are available?

The risk of false positives is always present in empirical analysis and is magnified in our case by the size of data and number of models. Only a careful and honest statistical analysis can attenuate this risk. In particular, as mentioned, we suggest to compare alternative indicators and methods over a training sample, select the preferred approach or combine a few of them, and then test if they remain valid in a genuine (not previously used) sample.

Are correlations mistaken for causes when interpreting the results?

Again, this is a common problem in empirical analysis that is exacerbated in a Big Data context. For example, a large number of internet searches for “filing for unemployment" can predict future unemployment without, naturally, causing it. Hence, we should abstain from a causal interpretation of the results, unless it can be sustained by economic theory and/or institutional considerations.

Is there instability in the nowcasting ability of specific Big Data based indicators?

Instability is a common source of errors in nowcasting exercise also with standard indicators. It can be due to a variety of reasons, such as the financial crisis, more general institutional changes, discontinuity in data provision, etc.

In the case of Big Data, there are some additional specific reasons for instability, such as the increasing use of internet and what Lazer et al. (2014) labeled “Algorithm Dynamics", namely the continuous changes made by engineers to improve the commercial service and by consumers in using that service.

Instability is indeed often ignored in the current big data literature, while it is potentially relevant, as we know well from the economic forecasting literature. Unfortunately, detecting and curing instability is complex, even more so in a big data context. However, some fixes can be tried, borrowing from the recent econometric literature on handling structural breaks.

Recommendations

- Conducting an in-depth real-time or pseudo real-time simulation of competing models in order to evaluate their relative performance in nowcasting the variable of interest.
- Models including Big Data should be preferred when they significantly lead to an improvement of the reliability and accuracy of the nowcasting at the same point in time.
- Models including Big Data should also be preferred when they allow for timelier nowcasting without any significant loss in terms of reliability and accuracy.
- The selected nowcasting models with Big Data should not suffer of the drawbacks discussed in the description of this step or they should minimise their impact so that the final estimates are not significantly affected.

17.2.7 Step 7: Implementation of big data based nowcasting

Description

In case the in-depth comparative analysis carried out in the previous steps suggests that the use of Big Data can improve the nowcasting precision or can lead to timelier ones for a given variable of interest, they can be then implemented. At this stage, the institution in charge of producing nowcasts should take several

Table 17.1: Proposed steps to design a sequential approach

Steps	Title	Aim
Step 1	Big Data usefulness within a nowcasting exercise	Checking for the existence of adequate Big Data sources
Step 2	Big Data search	Identification of the appropriate Big Data
Step 3	Assessment of big-data accessibility and quality	Verification of Big Data availability and quality including the presence of any bias
Step 4	Big data preparation	Move from unstructured to structured Big Data and filtering out undesirable effects
Step 5	Designing a Big Data modelling strategy	Choice of proper econometric methods and specifications
Step 6	Results evaluation of Big Data based nowcasting	Checking for the effective contribution of Big Data
Step 7	Implementation of big data based nowcasting	Timing and scheduling for the new nowcasting

relevant decisions related to the number of the nowcasting rounds to be implemented and their scheduling. For example, it is possible to decide to publish just one nowcast (e.g. at the very end of the reference period or at the very beginning of the following one), to produce two nowcasts (e.g. one in the middle of the reference period and one at the very end), or to produce a sequence of nowcasts scheduled at weekly or even daily frequency. Such decisions should take into account, among other, the trade-off between timeliness and reliability, the user needs as well as some more institutional considerations. New Big Data based nowcasting should then be appropriately documented making them as transparent as possible for the users. Such new nowcasting should be pre-announced and the most appropriate dissemination strategy should be identified in order to avoid any kind of misinterpretations by the users.

Recommendations

- Implementing and publishing the most reliable nowcasts available either at the end of the reference period or at the beginning of the following one.
- Moving towards a daily or weekly update on nowcasting already during the reference period, only after detailed pros and cons analysis and a consultation of the most relevant stakeholders.
- The new Big Data based nowcasting should be accompanied by clear metadata and widely available reference and methodological papers.

17.3 Conclusions

This chapter presents an operational step by step approach for using Big Data in a nowcasting exercise. It aims to facilitate the activity of experts involved in the construction of nowcasting by providing a set of recommendations associated to various operational steps. All the steps are summarized in the Table below.

The first step should be an a priori assessment of the potential usefulness of Big Data for a specific indicator of interest, such as GDP growth, inflation or unemployment. This requires to evaluate the quality of the existing nowcasts and whether any identified problems, such as bias or inefficiency or large errors in specific periods, can be fixed by adding information as potentially available in Big Data based indicators. Similarly, it should be

considered whether these additional indicators could improve the timeliness, frequency of release and extent of revision of the nowcasts. Relevant information can be gathered by looking at existing empirical studies focusing on similar variables or countries.

Once Big Data passes the "need check" in the preliminary step, the second step of the Big Data based nowcasting exercise is a careful search for the specific Big Data to be collected. As we have seen, there are many potential providers, which can be grouped into Social Networks, Traditional Business Systems, and the Internet of Things. Naturally, it is not possible to give general guidelines on a preferred data source, as its choice is heavily dependent on the target indicator of the nowcasting exercise.

Having identified the preferred source of Big Data, the third step requires to assess the availability and quality of the data. A relevant issue is whether direct data collection is needed, which can be very costly, or a provider makes the data available. In case a provider is available, its reliability (and cost) should be assessed, together with the availability of meta data, the likelihood that continuity of data provision is guaranteed, and the possibility of customization (e.g., make the data available at higher frequency, with a particular disaggregation, for a longer sample, etc.). All these aspects are particularly relevant in the context of applications in official statistical offices. As the specific goal is nowcasting, it should be also carefully checked that the temporal dimension of the Big Data is long and homogeneous enough to allow for proper model estimation and evaluation of the resulting nowcasts.

The fourth step analyzes specific features of the collected Big Data. A first issue that is sometimes neglected is the amount of the required storage space and the associated need of specific hardware and software for storing and handling the Big Data. A second issue is the type of the Big Data, as it is often unstructured and may require a transformation into cross-sectional or time series observations. Even when already available in numerical format, pre-treatment of the Big Data is often needed to remove deterministic patterns and deal with data irregularities, such as outliers and missing observations.

The fifth step when nowcasting with Big Data requires to select the proper econometric technique. Here, it is important to be systematic about the correspondence between the nature of the Big Data setting and use under investigation and the method that is used. There is a number of dimensions along which we wish to differentiate.

- The first choice is between the use of methods suited for large but not huge datasets, and therefore applied to summaries of the Big Data (such as Google Trends, commonly used in nowcasting applications), or of techniques specifically designed for Big Data. Within each class of techniques there are then many options. For example, nowcasting with large datasets can be based on factor models, large BVARs, or shrinkage regressions. Huge datasets can be handled by sparse principal components, linear models combined with heuristic optimization,or a variety of machine learning methods (which, though, are generally developed assuming i.i.d. variables). It is difficult to provide an a priori ranking of all these techniques and there are few empirical comparisons and even fewer in a nowcasting context, so that is may be appropriate to apply and compare a few of them for nowcasting the specific indicator of interest.
- A second dimension is the frequency of the available data. If this frequency is mixed then specific techniques for mixed frequency data become relevant. Chief among them is unrestricted MIDAS which provides a very flexible framework of analysis and can be adapted to work together with most if not all Big Data methods be they machine learning of econometric. Of course if data are fully unstructured and have no time series format then the analysis becomes relevant and should be considered and applied.
- Yet another dimension relates to the purpose for which large datasets are considered. Possibilities include model or indicator selection, forecasting or a more structural analysis. In this case of course each purpose is best served by different methods and the choice of method crucially depends on the purpose. Most methods can be used for forecasting and so the choice has to be case dependent. We recommend that as many methods are possible are evaluated in a forecasting context although past experience suggests that factor analysis and shrinkage methods can be of great use. For model or

indicator selection penalised regression and the MT methods seem to be appropriate and also have been reported to have good potential. Finally, for more structural analysis it is clear that it is likely that huge datasets are more difficult to accommodate. In this case, system methods that analyse the whole or a large proportion of the available data simultaneously, seem necessary for a satisfactory analytical outcome. Bayesian VAR models stand out as an appropriate method in this context.

The sixth step consists of a critical and comprehensive assessment of the contribution of Big Data for nowcasting the indicator of interest. In order to avoid, or at least reduce the extent of, data and model snooping, a cross-validation approach should be followed, whereby various models and indicators are estimated over a first sample and they are selected and/or pooled according to their performance, but then the performance of the preferred approaches is re-evaluated over a second sample. This procedure provides a reliable assessment of the gains in terms of enhanced nowcasting performance from the use of Big Data.

The final step is the implementation of the new Big Data based nowcasting. This requires to take several relevant decisions related to the number of the nowcasting rounds to be implemented and their scheduling. For example, it is possible to decide to publish just one nowcast (e.g. at the very end of the reference period or at the very beginning of the following one), to produce two nowcasts (e.g. one in the middle of the reference period and one at the very end), or to produce a sequence of nowcasts scheduled at weekly or even daily frequency. Such decisions should take into account, among other, the trade-off between timeliness and reliability, the user needs as well as some more institutional considerations. The most appropriate dissemination strategy should be also identified in order to avoid any kind of misinterpretations by the users.

To conclude, we are very confident that Big Data are precious also in a nowcasting context, not only to reduce the errors but also to improve the timeliness, frequency of release and extent of revision. We hope that the approach we have developed in this project will be useful.

Bibliography

Harford, T. (2014, April) Big data: Are we making a big mistake? Financial Times. Retrieved from http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html

Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). "The Parable of Google Flu: Traps in Big Data Analysis", *Science*, 143, 1203-1205.

VIII Constructing and Evaluating Rapid Estimates

18 Evaluating Rapid Estimates

Authors

Dominique Ladiray
INSEE
dominique.ladiray@insee.fr

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

18.1 Introduction

Throughout the handbook, we have presented a variety of techniques accompanied by some advices and recommendations for the construction of a rapid estimate model. In this chapter we will focus on the quality assessment, i.e. the statistical properties, and the validation, i.e. the quality of the estimates, of the selected model.

According to the European and other international standards, guidelines and good practices, "... the credibility of the statistics is enhanced by ... a sound methodology, appropriate statistical procedures, ...", Eurostat (2011). In this respect, the choice of the model(s) results from a global procedure including on one hand economic knowledge and reasoning, and on the other hand, outcomes of static and dynamic simulations. The chapter, which is somehow a summary of the previous technical chapters[1], is organized according to the properties a rapid estimate model should present:

- The model should be interpretable (Section 18.2),
- The model should be specified correctly (Section 18.3),
- The model should be stable and give unbiased forecasts (Section 18.4),
- The model should give good forecasts (Section 18.5).

Section 18.6 will advocate for the use of several different models at the same time in the production of a rapid estimate and Section 18.7 will briefly conclude the chapter.

18.2 The model should be interpretable

The first quality a rapid estimate model must fulfill is to be trusted and accepted by users.

- To be easily acceptable by a statistical office, a model should be explicable and understandable from the economic point of view. Even if "experience does not indicate that one should expect economic laws to be simple in nature", Granger (1969), the link between the target variable and the explanatory variables should have a clear economic interpretation and should ideally be persistent over the time. In particular, it is recommended to avoid in the model curious variables as well as strange and non interpretable lags and transformation, even for sensible variables.
- But on the other hand, the model should not be influenced by one economic theory or another. For example, compilers of rapid estimates should not nowcast the inflation rate using the output gap as an explanatory variable since it implicitly means the acceptance the hypothesis of Okun's law and Philips curve.
- It is also important to note that a meaningful model from the economic point of view is likely to be more stable or persistent and more manageable in production.

18.3 The model should be specified correctly

18.3.1 Timeliness

The timeliness of the model should be clearly stated and in agreement with user's needs.

- The model should be specified for a specific time lag or delay with respect to the reference period. In other words, the model should deliver estimates "d" days after the end of the reference period. For example, as the Flash Estimate EU and euro area quarterly GDP is delivered 30 days after the end of

[1] There is no formula in this chapter as most of the mentioned indicators have already been defined in previous chapters.

the reference quarter (at $t + 30$), a useful rapid estimate should be of course delivered before, at $t + 0$ or $t + 15$ for example.

- Of course, the rapid estimate model should take into account all quarterly or monthly information available at this time.

18.3.2 Statistical tests

The model must also verify some statistical criteria.

- Stationarity tests: Most of the methods used in the construction of rapid estimates require stationarity and it is important to check this point. It is for example better to work with growth rates, that are more likely stationary, than with levels. Many stationarity tests, and perhaps too many, are usually implemented in statistical packages that can be used: augmented Dickey-Fuller unit root and Engle-Granger cointegration Tests, Phillips-Perron unit root and cointegration tests, ERS and Ng-Perron unit root tests, Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) unit root test and Shin cointegration test etc.
- Misspecification tests: Omitting an important variable in the model or specifying a wrong functional form are model misspecifications which will affect the model performances. For example, a functional form misspecification generally means that the model does not account for some important nonlinearities and causes bias in the remaining parameter estimators. Ramsey (1969) proposed a general functional form misspecification test (Regression specification Error Test, RESET tests), to check whether power transforms need to be added to a model and which has proven to be useful. Other tests have been developed in the case of non-nested alternatives, Davidson and MacKinnon (1981), or more complicated cases, see for example Wooldridge (1994).
- Goodness-of-fit, Parsimony and other points to check The following usual check-list must be completed using the adhoc tests or measures, usually available in standard econometric packages.
  - The model should be able to produce forecasts (the roots of the auto-regressive part should be outside the unit circle);
  - All coefficients should be significant;
  - Residual should have no significant structure (white noise is an ideal but unexpected situation);
  - The R-square, or other proximity criteria like the MAPE, should be "correct"[2];
  - Possible colinearity problems must be checked and fixed;
  - The model should be parsimonious (AIC, BIC or other statistics taking into account the number of parameters and the number of observations should be used).

18.4 The model should be stable and give unbiased forecasts

It is clear that to be trusted, a model should be as stable as possible. This should not be taken for granted as breaks are present in many macroeconomic time series (e.g. Stock and Watson (1996) and Ang and Bekaert (2002)) and one of the major reasons of poor predictive performance (e.g. Clements and Hendry (1998)).

A large literature exists on the subject, proposing univariate forecasting methods which are robust to breaks (e.g. Pesaran and Timmermann (2007), Eklund et al. (2013)) or which formally model the break process (e.g. Pesaran et al. (2006), Koop and Potter (2007) and Maheu and Gordon (2008)).

[2] There is no precise threshold above which a R-square should be good. The R-square value must be appreciated simultaneously with other criteria. For example, one could prefer a model more parsimonious, or that can be explained more easily, even if the R-square is smaller.

It is still an open question to find the best method or model taking into account structural breaks. But in many cases, the formal modeling of the break process is important in achieving good forecast performance and there are many cases where simple, rolling window based forecasts perform well (Bauwens et al. (2011)).

The stability of the model and of the parameters can be assessed using several well-known tests using sub-samples, rolling windows or recursive estimation: Chow test (Chow (1960)), Bai-Perron test (Bai and Perron (1998)), CUSUM tests (Brown et al. (1975)) etc.

Statistics on the size of the one-step ahead forecast error also provide a first assessment of the forecast quality and the potential bias in the forecasts. The Root Mean Square Forecast Error (RMSFE) is a good synthesis.

18.5 The model should give good forecasts

18.5.1 Real-time estimation

The evaluation of the quality of the rapid estimates based on the last available vintage of the target variable, even by means of recursive estimation, is usually too optimistic and the performances of a model should ideally be evaluated in real-time. More and more real-time databases are available nowadays and they can be used for this purpose.

This is certainly the best way to get a clear assessment of the model performances. Using the real data available at a past dates allows forecasting the value for date t+1 and checking the accuracy of the model in "real production time".

Some descriptive statistics can be used:

- The mean of the forecast errors will indicate if the model is biased or not;
- The RMSFE computed on the one-step ahead forecast errors and the coherence between the growth rate signs are two very good indicators to summarize the quality of the model, and they can also be used to compare competitive models.

18.5.2 Comparing 2 models: the Diebold-Mariano statistics

Diebold and Mariano (1995) have tested two competitive models in order to find the better one in term of forecasting powerfulness. They evaluate the null hypothesis of equal predictive ability between two models. In particular, their statistic uses the discrepancies in forecast errors (available for all the forecast range) between the two competitive models.

A revised Diebold and Mariano statistic was proposed by Harvey et al. (1997). Under the null hypothesis of equal predictive ability, this revised test statistic is distributed as a t-statistic with $N1$ degrees of freedom. In that way, in the case of null hypothesis rejection, the different predictive ability of these models would be found. Sometimes the null hypothesis could not be tested (the Diebold and Mariano statistic cannot be calculated). In such cases, the RMSFE measure is not able to give sufficient information to discriminate among compared models, so that it is impossible to identify the best one. Nevertheless it is still the mostly widely performed measure according to some indications provided by researchers. In particular it must be specified how the RMSFE should be used in comparing different competitive models. Furthermore the first drawback of the RMSFE is the lack of invariance with respect to linear transformation of the forecasting model.

Another drawback is the non-invariance of the RMSFE in the case of multi-step forecasts. Both the previous mentioned problems are well discussed in Clements and Hendry (1998). The generalized forecast-error second-moment (GFESM), which is the determinant of the forecast-error second moment matrix represents a solution of the lack of invariance problems (see Clements and Hendry (1998)).

18.5.3 Density forecasts

It has become increasingly well understood that the main question which both producers an users are confronted to is not if a given type of rapid estimate is proving to be 'right' or 'wrong' but how large is the associated degree of uncertainty. Point estimates, which appear to be the traditional way of producing and disseminating rapid estimates, are better seen as the central points of ranges of uncertainty associated to each rapid estimate. Obviously this is not true when the empirical density is not symmetric as the theoretical one is supposed to be. As an example suppose that a nowcast or flash estimate model of the GDP growth returns a value of say, 2%. In this situation we can suppose that users should not be surprised if the actual growth, represented by the first regular estimate of the GDP growth, turns out to be a little higher or lower than the estimated one. Furthermore , perhaps, in times of heightened economic uncertainty, like close to a pick or during a recessionary phase„ they should not be very surprised if it turns out to be much higher or lower. Consequently, to provide a complete description of the forecasting ability of a rapid estimate model we need to complement the point estimates with some measures of uncertainty associated to them. Nowadays several forecasters produce density nowcasts and/or forecasts, which have been popularized by means of the so called "fan charts" firstly introduced by the Bank of England.

More formally, density forecasts of GDP growth, say, provide an estimate of the empirical probability distribution of its possible future values. In contrast to interval forecasts, which give the probability that the outcome will fall within a given interval, such as GDP growth falling within its target range, density forecasts provide a complete description of the uncertainty associated with a forecast. They can thus be seen to provide information on all possible intervals.

The ways density forecasts are constructed are detailed in Chapter 12 of this handbook.

18.6 The model or the model(s)?

When constructing rapid estimates it might be dangerous, at least in some cases, to rely on a single model, although carefully selected. This is not only true in presence of structural changes but, due to a number of reasons and factors, also in relatively normal period. Some reasons for which a single model should be not considered the best solution for producing rapid estimates are listed below:

- When constructing a model, it is always a good practice to compare it to a "benchmarking model". The Random Walk model, the most simple and often difficult to beat model, should be considered. If your model does not give better results, why should you trust it?
- During production time, it could occur that a variable is not available when needed so that the availability of an alternative model may be very useful.
- Some problems may occur when fitting the model. Sometimes a particular model may not be appropriate over a specific period of time. For this reason Blake et al. (2000) and Camba-Mendez et al. (2001) introduce the possibility to recursively choose the best fitting model, among a chosen subset, at each period in the forecasting evaluation period.
- The first estimates of European aggregates must often be computed by Eurostat from partial information. The practice of the European institute varies according to the indicator. If a single model is used for the GDP flash estimate, several competing models are used for the HICP.

In statistics, Bayesian Model Averaging (BMA) offers a conceptually elegant mean of dealing with "model uncertainty". BMA forecasts condition not on a single "best" model but take a weighted average over a range of candidate models; see Hoeting et al. (1999). This follows from the appreciation of the fact that, although one model may be "better" than the others, we may not select it with probability one. We may not be sure that it always returns the best forecast. Therefore, if we consider this single forecast alone, we would be overstating its precision.

Similarly in practical macroeconomic forecasting exercises, whether within a Bayesian context or not, it is a stylized fact that combination forecasts are hard to beat. The estimated parameters of a single forecasting model are commonly found to exhibit instabilities and these can be difficult to identify in real-time. In the presence of these so-called “uncertain instabilities” it can be helpful to combine the evidence from many models. For example, Clark and McCracken (2009) examine the scope for taking linear combinations of point forecasts in real time, motivated by the aim of circumventing the uncertain instabilities in any particular specification. In a series of influential papers, Stock and Watson (2004) have documented the robust performance of point forecast combinations using various types of models for numerous economic and financial variables. Selecting a single model has little appeal under ’uncertain instabilities’ when the single best model suffers from instability. This might happen either if the “true” model is not within the model space considered by the modeler, or if the model selection process performs poorely on short macroeconomic samples. We may better approximate the truth, and account for the uncertainty in model selection, by combining forecasts.

Methods for combining point forecasts and examples can be found in Hall and Mitchell (2007) and Mazzi et al. (2014).

18.7 Conclusions

The evaluation of a model producing rapid estimates is therefore a trade-off between pragmatism and technique, between statistical tests and economic interpretation. We have shown how many aspects have to be taken into account when evaluating and comparing rapid estimates models. We have to admit that not all recommended procedures described in the chapter are easily applicable by statistical institutions so that the recommendation is to go step by step firstly considering the most evident and easily implementable ones then to go further. For example it is not so difficult to accept and implement principles related to the absence of theoretical influence in the model or to the correct specification of the model and its forecasting accuracy. A bit more complicate could be to adopt the real-time investigation and comparison of rapid estimates models especially if statistical institutions do not collect and store real-time vintages. Nevertheless, due to the importance of this point for the evaluation of rapid estimate models, this chapter could help in sensitizing statistical authorities to this point. Concerning the presentation of rapid estimates as probabilistic density this is a point to be carefully evaluated and a decision should be taken also considering the risk of confusing users or to lower the intrinsic communication power that a single number has.

Bibliography

Ang A. and Bekaert G. (2002). Regime switches in interest rates, Journal of Business and Economic Statistics, 20, 163-182.

Bai J. and Perron P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes, Econometrica, 66, 47-78.

Bauwens L., Koop G., Korobilis D. and Rombouts J.V.K. (2011). The Contribution of Structural Break Models to Forecasting Macroeconomic Series, Working Paper WP 11-38, The Rimini Centre for Economic Analysis.

Blake A., Kapetanios G. and Weale M. (2000). Industrial Production Nowcasting. Report submitted to Eurostat by National Institute of Economic and Social Research.

Brown R.L., Durbin J. and Evans J.M. (1975). Techniques for Testing the Constancy of Regression Relationships over Time, Journal of the Royal Statistical Society (B) 37:2, 149-163.

Camba-Mendez G., Kapetanios G., Smith R. and Weale M. (2001). An Automatic Leading Indicator of Economic Activity: Forecasting GDP Growth for European Countries, Econometrics Journal, 4 (1), S56-S90.

Chow G. (1960). Tests of Equality between Sets of Coefficients in Two Linear Regressions, Econometrica, 28, 531-534.

Clark T.E. and McCracken M.W. (2009). Averaging forecasts from VARs with uncertain instabilities, Journal of Applied Econometrics, 25 (1), 5-29.

Eklund J., Kapetanios G. and Price S. (2013). Robust forecast methods and monitoring during structural change, Manchester School, Vol. 81, S3 pages 3-27.

Eurostat (2011). European Statistics Code of Practice - revised edition 2011, available at http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955.

Clements M.P. and Hendry D.F. (1998). Forecasting Economic Time Series, Cambridge University Press.

Davidson R. and MacKinnon J.G. (1981). Several tests for model specification in the presence of alternative hypotheses, Econometrica 49, 781-793.

Diebold F.X. and Mariano R. (1995). Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, 253-265.

Giraitis L., Kapetanios G. and Price S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change, Journal of Econometrics, Vol. 177, pages 153-170, 2013.

Granger C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica, Vol. 37, No. 3. (Aug., 1969), pp. 424-438.

Hall, S. G. and Mitchell, J. (2007), Combining density forecasts, International Journal of Forecasting 23, 1-13.

Harvey D.I., Leybourne S.J. and Newbold P. (1997). Testing the Equality of Prediction Mean Square Errors, International Journal of Forecasting, 13, 273-281.

Hoeting J.A., Madigan D., Raftery A.E. and Volinsky C.T. (1999). Bayesian model averaging: a tutorial, Statistical Science 14, 382-417.

Koop G. and Potter S. (2007). Estimation and forecasting with multiple breaks, Review of Economic Studies, 74, 763-789.

Maheu J. and Gordon S. (2008). Learning, forecasting and structural breaks, Journal of Applied Econometrics, 23, 553-583.

Mazzi, G-L. and Mitchell, J. and Montana, G. (2014), Density nowcasts and model combination: nowcasting Euro-area GDP growth over the 2008-9 recession, Oxford Bulletin of Economics and Statistics, Vol. 76, pp. 233âĂŞ256.

Pesaran M.H., Pettenuzzo D. and Timmermann A. (2006). Forecasting time series subject to multiple structural breaks, Review of Economic Studies, 73, 1057-1084.

Pesaran M.H. and Timmermann A. (2007). Selection of estimation window in the presence of breaks, Journal of Econometrics, 137, 134-161.

Ramsey J.B. (1969). Tests for specification errors in classical linear least-squares analysis, Journal of the Royal Statistical Society, Series B, 71, 350-371.

Stock J.H. and Watson M.W. (1996). Evidence on structural instability in macroeconomic time series relations, Journal of Business & Economic Statistics, 14, 11-30.

Stock J.H. and Watson M.W. (2004). Combination forecasts of output growth in a seven-country data set, Journal of Forecasting 23, 405-430.

Wooldridge J.M. (1994). A simple specification test for the predictive ability of transformation models, Review of Economics and Statistics 76, 59-65.

19 Guidelines for the Construction of Rapid Estimates

Authors

Gian Luigi Mazzi
Independent Expert
glmazzi@pt.lu

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

Table of Contents

Handbook on Rapid Estimtates
Edited by G. L. Mazzi, co-edited by D. Ladiray

19.1 Introduction

19.1.1 Motivation for the guidelines

These guidelines suggest the use of best practices in the construction of rapid estimates aiming at:

- developing a harmonised system of rapid estimates at national and regional level
- enhancing comparability of rapid estimates across countries and regions
- increasing the effectiveness of real-time monitoring of the macroeconomic situation.

These guidelines follow the taxonomy and definitions introduced in the Glossary annexed to chapter 2 (sections 2.A and 2.B) and aim to provide best practices for the main typologies of rapid estimates.

19.1.2 Scope of the guidelines

These guidelines are firstly addressed at statistical institutions already involved in the compilation of rapid estimates or evaluating the possibility to start working in this field. Topics covered in the guidelines and proposed recommendations make them also of high interest for public and private institutions working in compiling rapid estimates.

The guidelines are aimed at anyone whose work is related in a more or less strict way to the compilation of rapid estimates; they have been conceived both for experts and beginners since they also present a didactic orientation.

The guidelines cover all issues related to the construction, the regular production and the maintenance of rapid estimates: from data issues to compilation aspects of various kinds of rapid estimates, to quality and dissemination issues. As shown in the glossary (see Barcellan et al. (2009) and 2.B), there is a large variety of rapid estimates which can be defined by combining various characteristics in different ways. In the sake of brevity, in these guidelines we are concentrating our attention on three main types of rapid estimates:

- flash estimates to be produced after the end of the reference period by using as much as possible a methodology in line with the one used for the regular estimates;
- nowcasting to be produced right before or after the end of the reference period using statistical or econometric techniques;
- advanced estimates to be computed during the reference period by means of econometric and forecasting techniques.

These guidelines are restricted to macro-economic rapid estimates so that they do not cover socio or socio-economic cases even if some of the recommendations and suggestions provided here can be relatively easily translated to other domains.

The guidelines are based on a set of principles, presented in section 19.1.4, which give some general rules to be followed when compiling rapid estimates.

They are structured in seven sections, dealing with general and specific issues of rapid estimates such as data issues, compilation of rapid estimates etc. Each section is subdivided in a number of subsections (items) dealing with a specific aspect. All items are presented following a standardized 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, presents 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 prevent 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 some not recommended procedures.

The objective of the guidelines is helping producers of rapid estimates to move towards alternative (A). Careful considerations and, possibly, prompt measures should be taken whenever alternative (C) is in use.

19.1.3 Advantages and drawbacks

Implementing these practical guidelines will enhance the transparency and clarity of rapid estimates. It will also increase the cross-countries/regions comparability and will contribute to the development of an effective real-time monitoring system of the economic situation. Even though the implementation of these guidelines may require additional efforts (in both human and IT resources), the benefits will outweigh the initial efforts. The decision of not applying the guidelines will prevent at least partially the process of producing more and more comparable rapid estimates to facilitate a global view of the short-term economic situation.

19.1.4 Principles for the compilation of rapid estimates

The production of official statistics follows a rigorous scrutiny in terms of its adherence to internationally agreed principles as reflected in the Quality Assurance Framework of the European Statistical System (ESS Code of practice (2011))[1] and the principles of the United Nations template for a Generic National Quality Assurance Framework (NQAF Template (2012)) developed by the Expert Group on NQAF and endorsed by the United Nations Statistical Commission in 2012.[2]

Below are the principles that are particularly important for the compilation of rapid estimates and explained within the context of rapid estimates.

- **Impartiality and objectivity**

  Rapid estimates should deliver messages based on statistical evidence coming from data; they should be produced and disseminated independently from political or economic theories.

- **Methodological soundness**

  Rapid estimates should be complied according to sound, well tested, and internationally recognised methods.

- **Accessibility and clarity**

  Rapid estimates should be accompanied by a clear documentation easily accessible and understandable by non-expert users. More detailed and complete documentation for expert users should also be available publicly or on request.

- **Transparency**

  All steps of the compilation process of rapid estimates should be clearly listed, described and documented.

[1] See http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf
[2] See https://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx

- **Interpretability and readability**

  The dissemination of rapid estimates should focus on an easily interpretable message for all users; the use of visualisation and graphical tools is recommended.

- **Coherence and Comparability**

  Various kinds of rapid estimates compiled by the same institution with different timing should deliver consistent messages. Possible inconsistency, when happening, should be highlighted, explained and, whenever possible, correcting measure should be undertaken. Consistency should be regularly monitored .All institutions involved in the compilation of rapid estimates should foster the comparability of their estimates at national level, across countries and regions.

19.2 General aspects

19.2.1 A general policy for the compilation of rapid estimates

Description

A general policy for the compilation of rapid estimates should be defined on the base of the set of principles presented in the annex above. It should contain information related at least to: the typology of rapid estimates to be compiled, the adopted methodological and quality framework, the release and revision strategy for such estimates, and dissemination aspects. The availability of such a policy would increase the transparency of the production process and the trust in the producer.

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

(A) Elaborate and disseminate a general policy for the compilation of rapid estimates, based on a set of agreed principles, and detailing all key aspect related to the production, release, revision, quality assessment and dissemination of such estimates;

(B) Elaborate and disseminate a policy for the compilation of rapid estimates, only partially in line with a set of agreed principles, covering at least partially key production, release, revision, quality assessment and dissemination aspects for such estimates;

(C) Adopt an incomplete policy or a policy not in line with agreed principles, or lack of any policy.

19.3 Domain specific policies for rapid estimates

Description

The need for rapid estimates can depend upon the statistical domain in different ways and for several reasons, for example the timeliness of regular estimates could be considered already satisfactory by the users. Furthermore the characteristics of the process of building-up information can be very different by domain, as well as the degree of volatility of the target variables for which rapid estimates should eventually be compiled, the relevance of the variables and the available resources, and the information used as input to the compilation process.

As a consequence, specific policies for rapid estimates should be adopted at domain level in order to incorporate data characteristics and other peculiarities.

Options

- Develop domain specific policies for the compilation of rapid estimates fully complaint with the general one;
- Derive domain specific policies for rapid estimates only partially compliant with the general policy;
- Compile domain specific policies for rapid estimates even partially inconsistent with the general policy;
- Not developing any domain specific policy for rapid estimates.

Alternatives

(A) Develop domain specific policies for rapid estimates fully compliant with the general policy and taking into account all issues specific to each statistical domain;

(B) Develop domain specific policies for rapid estimates only partially compliant with the general policy or covering only partially issues specific to each statistical domain;

(C) Develop domain specific policies for rapid estimates inconsistent with the general policy or do not develop at all any domain specific policy.

19.3.1 Stability of general and domain specific policies

Description

Keeping general and domain specific policies for compiling rapid estimates stable over time will increase users' confidence and will allow producers to work within a stable framework. On the other hand, assuming that the general and domain specific policies should never be revised is not realistic since the economic and statistical conditions will evolve, as well as the relevance of a particular domain and consequently policies would need to be updated. Institutions should then find the right balance between keeping policies as much as possible stable and ensuring that they are in line with an evolving world.

Options

- Keep general and specific policies for rapid estimates as much as possible stable over time and change them only after a thorough investigation, pre-announcing changes;
- Revise general and specific policies on the base of a regular calendar, for example every five years, in order to ensure sufficient stability;
- Revise policies often, for example every year;
- Revise policies for rapid estimates on irregular base and without any pre-announcement;
- Never revise rapid estimate policies;

Alternatives

(A) Keep the general and the domain specific policies for the compilation of rapid estimates as much as possible stable over time and update them only when significant changes occur either in the economic system or in the statistical system; changes to the general and domain specific policies should be pre-announced;

(B) Update general and domain specific policies on a fixed scheduling (for example every 5 years) in order to ensure sufficient stability; changes should be pre-announced;

(C) Update general and domain specific policies too often on an irregular base and without any pre-announcement; never update the general and the domain specific policies.

19.3.2 Quality assurance framework for rapid estimates

Description

Quality assessment of rapid estimates needs to consider all five dimensions of statistical output quality, as listed in the European Statistics Code of Practice ESS Code of practice (2011) and in the United Nations template for a Generic National Quality Assurance Framework NQAF Template (2012):

- Relevance
- Accuracy and reliability
- Timeliness and punctuality
- Coherence and compatibility
- Accessibility and clarity

Measures identified for each dimension can be qualitative or quantitative — where qualitative measures will normally have a "Yes" or "No" value, while quantitative measures will normally be test statistics with the direct interpretation of "Pass" or "Fail". For example, relevance could be measurable qualitatively through consultation with users, while accuracy and reliability are generally measured quantitatively.

Since rapid estimates might be, at least in some cases, somewhat different from traditional official statistical products, the quality framework needs to be adapted to incorporate their specificities without changing its grounds. Two quality dimensions particularly relevant for rapid estimates are timeliness and accuracy (see chapter 4).

Specific aspects linked to the assessment of the quality of a rapid estimate, for example if its value is supposed to anticipate the subsequent release or the final estimate, and if the focus is more on the timeliness of the results, or their accuracy, or both should be taken in consideration when assessing its quality.

Options

- Use the quality framework for official statistics duly amended to incorporate specific features of rapid estimates, in particular concerning the timeliness and the accuracy dimensions;
- Use the default quality framework for official statistics;
- Use an ad-hoc quality framework for rapid estimates;
- Do not use any quality framework.

Alternatives

(A) Use the standard quality framework for official statistics duly amended to take into account all specificities of rapid estimates, in particular concerning the timeliness and the accuracy dimensions;

(B) Use the default quality framework for official statistics;

(C) Use an ad-hoc quality framework for rapid estimates or not use any quality framework.

19.4 Data issues

19.4.1 Input data availability and quality

Description

The availability of a large set of macro-economic statistical indicators produced at infra-annual frequencies (possibly monthly) is a crucial pre-requirement for the compilation of rapid estimates, especially nowcasting and advanced ones. Flash estimates can also benefit of a large input dataset, especially in order to fill gaps or identify proxies to replace totally unavailable indicators. These macroeconomic statistical indicators should meet some important requirements:

- Availability of a sufficiently long time series (not less than 10 years);
- Regular behaviour with limited number of outliers, structural changes and other irregularities;
- Compiled according to international agreed standards;
- Regularly produced and disseminated on the basis of a clear release calendar;
- Accompanied by a clear documentation;
- Having coincident or leading behaviour with respect to a small number of target variables for which rapid estimates are compiled.

The compilers of rapid estimates should be able to regularly monitor the overall quality of the statistical indicators they are going to use as input to the computation of rapid estimates. This monitoring will allow them to have the best possible knowledge of data, of their generating process as well as of their advantages and drawbacks.

Options

- Develop a small-size database containing relevant macro-economic statistical indicators available at sub-annual frequencies, and implement on it an automatic system of quality monitoring to be periodically run, for example every week or month;
- Regularly check the quality of macro-economic statistical indicators using monitoring systems already in place, if any, or conducting a non-automatic and non-exhaustive assessment;
- Performing an assessment of macro-economic statistical indicators irregularly;
- Do not perform a specific quality assessment.

Alternatives

(A) Develop and maintain a small-size database for macro-economic statistical indicators potentially relevant for the compilation of rapid estimates and implement on it an automatic system of quality monitoring and assessment to be run on a regular basis (ideally every month);

(B) Follow up the quality of relevant macro-economic statistical indicators by using already available quality information and complement it by additional checks conducted on a regular basis but in a non-automatic way;

(C) Perform a non-regular quality control not necessarily covering all relevant statistical indicators or do not perform any quality control at all.

19.4.2 Real-time vintages

The construction of rapid estimates could considerably benefit from the availability of real-time vintages of the most relevant macro-economic statistical indicators. Historical vintages will allow compilers of rapid estimates to conduct a comparative simulation in real-time over the past to assess the performance of alternative models in order to select the best performing one. Furthermore, they will permit to evaluate the loss of accuracy associated to different timeliness in real-time.

Developing and maintaining a system for an automatic collection and storage of vintages is not an easy task and it implies a certain investment on ICT and human resources by statistical institutes which are the best placed 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 country coverage remains still unsatisfactory.

It is worth to notice that the frequency of vintages should be chosen in accordance with the nature of the rapid estimate and input data availability; for example, when estimating a quarterly variable it could be superfluous to have daily vintages; when choosing the optimal timeliness in relation to data availability a thorough analysis of input data releases has to be performed to decide about the optimal frequency of vintages collection and storage.

Options

- Regularly collect and store vintages for most relevant macro-economic statistical indicators in an appropriate ICT framework already in place, ensuring that the coverage and frequency are appropriate to the production of the targeted rapid estimate;
- Develop a new system for collecting and storing vintages on a regular basis, with coverage and frequency appropriate to the production of the targeted rapid estimate;
- Collect vintages on a purely unplanned basis without having an appropriate storage system;
- Use national vintages already collected by other institutions;
- Do not collect or plan to collect any vintage.

Alternatives

(A) Use an already existing system for data-vintages collection and storing on regular basis or develop it to safeguard vintages availability for the future, ensuring that the coverage and frequency are appropriate to the targeted rapid estimate;

(B) Collect vintages on an unplanned basis without a good infrastructure for their storage or using already existing vintages collected by other institutions, provided that coverage and frequency are appropriate enough to the targeted rapid estimate;

(C) Do not collect or plan to collect any data-vintages.

19.4.3 Variable selection

Description

Starting from a set of macro-economic infra-annual statistical indicators, a variable selection exercise should identify the most suitable candidate series for the construction of rapid estimates. The first selection criterion is the ability of the series to accurately reproduce, anticipate or estimate in real-time the short-term patterns of the reference or target variables. This process should also allow identifying the leading/lagging structure of the selected variables.

The relevance of this step depends on the kind of rapid estimate: for flash estimates the information set is supposed to be very similar to the one used for regular estimates; selected variables should then help in filling gaps in some components of the information set or in identifying proxies for unavailable ones. For nowcasting and advanced estimates the selection step is essential to the construction of the information set used in the estimation process. This exercise can be performed by using parametric or non-parametric variable selection techniques, as well as graphical ones or techniques based on purely subjective appreciation and judgement.

Options

- Parametric variable selection techniques;
- Non-parametric variable selection techniques;
- Graphical techniques;
- Preselection based on personal and subjective data knowledge and/or appreciation;
- Combining several approaches;

Alternatives

(A) Combine the use of parametric and non-parametric techniques, complemented by graphical ones and expert knowledge to achieve the most efficient selection of variables to be used for rapid estimates;

(B) Rely either on parametric or not -parametric variable selection techniques only;

(C) Rely on graphical techniques or on subjective appreciation only.

19.5 Flash estimates

19.5.1 The dataset

Description

According to the definition (see Barcellan et al. (2009) and 2.B), flash estimates should be built up by using an information set as much possible similar to the one used for regular estimates. Unfortunately, anticipating the estimation time will imply that several observations could be missing or some indicators totally unavailable. Consequently, the original dataset should be amended in order to incorporate other variables allowing to estimate missing observations or to act as proxies for other series. In this context the role of soft data, mainly from economic tendency surveys, will be very relevant both for their timeliness and for their ability to replicate the pattern of hard data traditionally available much later.

Options

- Complement the dataset used for regular estimates with all indicators necessary to the estimation of missing data and to the replacement of missing indicators;
- Keep the dataset unchanged and estimate missing data by purely univariate techniques;
- Construct a dataset not fully reflecting the structure of the one used for regular estimates;
- Use a completely different dataset;

Alternatives

(A) Enlarge the information set used for regular estimates with all relevant indicators which could help to estimate missing values or to replace missing indicators. In the enlargement process, privilege tendency surveys and exclude financial indicators;

(B) Use the same dataset used for regular estimates replacing missing data by univariate forecasts;

(C) Use partially or totally different information sets from the one used for regular estimates.

19.5.2 The timing

Description

According to the definition (see Barcellan et al. (2009) and 2.B), flash estimates should be produced not later than $\frac{t}{2}$ days after the end of the reference period t, meaning 15 days for monthly data and 45 days for quarterly ones. When deciding to publish a flash estimate, statistical authorities should identify a specific point in time between the end of the reference period and $t + \frac{t}{2}$ days which maximizes the timeliness gain and minimize the loss of accuracy. The gain in terms of timeliness should be significant, since producing flash estimates only a few days in advance of regular ones would be useless and can maybe be dangerous for the credibility of the institution and miss-leading due to the publication of two different values within a very short time interval.

On the other hand, the loss of accuracy should not be so significant to compromise users' trust and the credibility of the institution.

Options

- Analyse in real-time the process of information acquisition and evaluate at various points in time the loss of accuracy with respect to more or less relevant gain in timeliness;
- Analyse the loss of information associated to various points in time by using final vintages;
- Identify the timing for the flash estimates based on users' needs and requests;
- Use subjective judgement and appreciation.

Alternatives

(A) Analyse in real-time the process of acquisition of information and conduct a real-time exercise to evaluate the possible loss of accuracy associated to different increase of timeliness, and select the one ensuring an optimal balance between timeliness improvement and accuracy loss;

(B) Analyse the process of acquisition of information using final vintages only and conduct an exercise to evaluate the possible loss of accuracy associated to different increases of timeliness;

(C) Identify the timing for the flash estimates using non statistical considerations only such as users' needs or personal appreciation.

19.5.3 Methodology

According to the definition (see Barcellan et al. (2009) and 2.B), flash estimates should be based on a statistical methodology as much as possible in line with the one used to produce regular estimates. For example, if the regular estimates are produced by using temporal disaggregation techniques, the same should happen for flash estimates. Obviously the processes cannot be identical; the regular production process generally involves a large number of series while for flash estimates only one of very few indicators are concerned. Furthermore, the amendment to the information set can require slightly modifications of the methodology. Finally the performance in producing flash estimates can also entail some changes to improve precision.

Options

- Use the same methodology as for regular estimates with no amendment but on a reduced scale;
- Analyse several variants of the methodology used for regular estimates;
- Consider models which tend to be substantially different from the one used for regular estimates but still pertaining to the same statistical field;
- Consider methods without any link with those used for regular estimates.

Alternatives

(A) Compare within a real-time exercise a number of variants of the methodology used to produce regular estimates, and chose the one which ensures the best improvement of timeliness with the minimum loss of accuracy. In case of very similar performances of variants, prefer the methodology closest to the one used for regular estimates;

(B) Compare a number of variants of the methodology used to produce regular estimates using final vintages, and chose the one which ensures the best improvement of timeliness with the minimum loss of accuracy. In case of very similar performances of variants, prefer the methodology closest to the one used for regular estimates. Or else, privilege the same methodology as for regular estimates but applied on a smaller scale;

(C) Use methods with weak or no link at all with the one used for regular estimates.

19.6 Nowcasting

19.6.1 The dataset

Description

Compilers of nowcasting (see Barcellan et al. (2009) and 2.B) less constrains in the construction of the reference dataset than those of flash estimates. They can select all candidate variables having a good forecasting power for the target variable, provided that they show a sound, statistically and economically explicable relationship with the target variable. The identification of candidate variables should be carried out either by using variable selection techniques or expert knowledge. Nowcasting information sets can be composed either by soft data only, meaning tendency surveys and eventually some financial indicators, or by hard and soft ones. Nowcasting can be built up using datasets of very different size, from a few variables to several hundreds; this can impact the statistical and econometric methods used.

Options

- Use a dataset composed by soft data only;
- Use a dataset composed by both hard and soft data;
- Use datasets of different size and composition;
- Restrict the dataset to variables showing a statistically and economically sound and interpretable relationship with the target variable;
- Use variable selection techniques together with experts knowledge to select variables.

Alternatives

(A) Compare within a real-time exercise the forecasting performance of datasets of various sizes and composition constructed combining variable selection techniques and expert knowledge. Select the best performing one and, in case of similar performance, privilege the smaller one. Exclude variables without a clear sound and interpretable relationship with the target variable;

(B) Conduct a smaller scale comparative exercise considering only small information sets of different composition and including only variables showing clear, sound and interpretable relationships with the target variable;

(C) Use a dataset based only on personal appreciation or use a dataset including variables with doubtful relationships with the target variable.

19.6.2 The timing

Description

According to the definition (see Barcellan et al. (2009) and 2.B), nowcasting should be released few days before or after the end of the reference period, assume between 5 days before and 5 days after the end of the reference period. Since in nowcasting soft data can play an important role, it is worth to note that opinion surveys are usually released right before the end of the month/quarter, while financial data usually right after. For this reason, one or two days of difference in the timing might significantly impact nowcasting accuracy.

Options

- Consider only release dates after the end of the reference period;
- Evaluate the timing by comparing the accuracy's change between alternative dates;
- Decide on the timing taking into account users' needs only;
- Select the timing based on subjective considerations;
- Always publish after the release of all relevant components of the dataset, but not later of few days after the end of the reference period.

Alternatives

(A) Select the best point in time within a real-time exercise in order to have the maximum increase of timeliness compatible with the minimum loss of accuracy;

(B) Publish nowcasting after the publication of the most relevant components of the dataset but not later than few days after the end of the reference period;

(C) Decide on the timing only on the base of subjective evaluations or users' needs.

19.6.3 The methodology

Description

From a theoretical point of view forecasting methods can be applied also for nowcasting exercises. In practical terms, especially when a statistical authority is producing nowcasting, it could be preferable to focus on relatively simple, robust and easily understandable models.

Looking at statistical methods, we can focus on three main categories of models which could be considered in nowcasting: dynamic regression models, vector autoregressive models in various specifications (VAR, SVAR, GVAR, VECM, . . .), and dynamic factor models.

Obviously, especially when the target variable is quarterly or even annual, the previously mentioned methods should be extended in order to deal with mixed frequency data or specific models such as the MIDAS or the unrestricted MIDAS. Finally nowcasting based on model averaging or forecasting combination should not be disregarded since they often provide very good results.

Options

- Compare all possible models to identify the best performing one;
- Concentrate on just one class of models such as dynamic factor models and disregard the others;
- Do not consider model combination or forecast combination for nowcasting;
- Use naive models or models not generally recognised in the forecasting literature.

Alternatives

(A) Compare various models and model specifications within a real-time exercise, incorporating also the ability of dealing with mixed frequency data if needed, and identify the best performing one(s). Compare nowcasting obtained by the best performing one(s) with those derived by model averaging and forecasting combination; select the model or model combination returning the best nowcasting. When the performance of a number of models is very similar, select the simplest one;

(B) Carry out a restricted simulation exercise either in real-time or on final vintages limited to a maximum of two classes of forecasting methods chosen on the base of expert evaluation and identify the best performing one(s). In case of similar performances chose the simplest model;

(C) Use naive models or models not internationally recognised based on subjective appreciation only.

19.7 Advanced estimates

19.7.1 The dataset

Description

In case of advanced estimates, meaning those rapid estimates produced during the reference period, the considerations about the dataset to be used are very similar to those ones of nowcasting. In particular, datasets for advanced estimates could be composed of soft data only or hard and soft data, and they could be of various size.

The main difference is that variables selected for the inclusion in a dataset used for constructing advanced estimates should have mainly leading characteristics with respect to the target variable. For this reason, particular attention in the variable selection process, independently by how it will be performed, should be paid to the identification and exclusion of possible spurious or unsound leading relationships.

Options

- Use a dataset composed by soft data only;
- Use a dataset composed by both hard and soft data;
- Use datasets of different size and composition;
- Restrict the dataset to variables showing a leading relationship with the target variable;
- Use variable selection techniques together with expert knowledge to select variables.

Alternatives

(A) Compare within a real-time exercise the forecasting performance of datasets of various sizes and composition constructed combining variable selection techniques and expert knowledge. Select the best performing one and, in case of similar performance, privilege the smaller one. Exclude variables with a spurious or unsound leading relationship with the target variable;

(B) Conduct a smaller scale comparative exercise considering only small information sets of different composition and including only variables showing sound leading relationship with the target variable;

(C) Use a dataset based only on personal appreciation or use a dataset including variables with spurious or unsound leading relationships with the target variable.

19.7.2 The timing

Description

According to the definition (see Barcellan et al. (2009) and 2.B), advanced estimates should be produced and released already during the reference period. Then, at least partially, they could conflict with nowcasting if produced and released just before the end of the reference period. For this reason, we assume here that advanced estimates are produced and released not later than few days before the end of the reference period.

Since advanced estimates are produced during the reference period, they are based on datasets which can be particularly incomplete and which can vary a lot over the time. This is particularly true when producing advanced estimates having as target variable an indicator with monthly frequency. The timing identification is then particularly important: timeliness is not any more an issue and the focus should be on minimising accuracy loss. It is worth to mention that statistical authorities should not anticipate too much advanced estimates in order to avoiding compromising their credibility and trustability.

Options

- Consider only release dates after a given amount of information becomes available;
- Evaluate the timing by comparing the accuracy's change between alternative dates;
- Decide on the timing taking into account users' needs only;
- Select the timing based on subjective considerations.

Alternatives

(A) Within a real-time exercise, select the best release timing after the acquisition of an acceptable amount of information which guaranties sufficient quality of advanced estimates;

(B) On the base of final vintages, select the best release timing after the acquisition of an acceptable amount of information which guaranties sufficient quality of advanced estimates;

(C) Decide on release timing only on the base of subjective evaluations or users' needs, without caring of estimates' quality.

19.7.3 The methodology

Description

Models used for advanced estimates are generally the same as those used for building up nowcasting, since both exercises can be viewed as special cases of forecasting. As for advanced estimates the correct identification and utilisation of the leading relationships linking variables in the dataset to the target one is of primary interest, very flexible models such as vector autoregressive and dynamic factor models should be preferred to dynamic regression ones which are more dependent on predefined lagging schemes.

Model choice and performance assessment should be carried out with particular carefulness and caution since the rapid evolution of available information can heavily impact the model choice and outcome. Model averaging and forecast combination could mitigate such negative impacts.

Options

- Compare all possible models to identify the best performing one;
- Concentrate on just one class of models such as dynamic factor models and disregard the others;
- Do not consider model combination or forecast combination for advanced estimates;
- Use naive models or models not generally recognised in the forecasting literature.

Alternatives

(A) Carry out a real-time simulation exercise using alternatives VARs and dynamic factor models specifications and identify the best performing one(s). Check whether or not the outcome of model averaging or forecasting combination outperforms that of the selected model(s). In case of similar performances chose the simplest model;

(B) Carry out a simulation exercise based on final vintages using alternatives VARs and dynamic factor models specifications and identify the best performing one(s). In case of similar performances chose the simplest model;

(C) Use models not able to deal with highly incomplete information sets and complex lagging structures or models not internationally recognised, based on subjective appreciation only.

19.8 Dissemination and presentation

19.8.1 Labelling of rapid estimates

Description

It is particularly important to clearly provide the definitions used in the compilation of rapid estimates when disseminating the data. The lack of consensus on the definitions of rapid estimates was recognized during the seminars organized by the United Nations and Eurostat in the period 2009-2010 which lead to the development of the glossary of rapid estimates by Eurostat Barcellan et al. (2009) (and 2.B). The adoption of the glossary of rapid estimates will facilitate the communication and increase the comparability across countries and regions.

Options

- Adopt the terminology proposed in the glossary;
- Use only partially the terminology proposed in the glossary;
- Use a national terminology not in line with that adopted in the glossary;
- Do not refer to any particular terminology.

Alternatives

(A) Use the terminology proposed in the glossary and label rapid estimates accordingly;

(B) Use only partially the glossary terminology;

(C) Use a terminology inconsistent with the glossary one or not use any particular terminology at all.

19.8.2 Data presentation

Description

The aim of rapid estimates is to estimate one additional observation when it cannot yet be computed within the regular production process. Consequently, the estimate could be considered as an additional point of a given time series and whether or not to append it to the official time series should be evaluated. This would make sense especially if the value of the additional point is computed by using a methodology sufficiently in line with the one used for the regular production, which is the case for flash estimates. However, when the methodology used is substantially different from the regular production one, as for nowcasting and advanced estimates, it might be preferable to present the estimate for the additional point in time separately.

Options

- Present flash estimates as an additional value of the time series vector describing the target variable, and nowcasting and advanced estimates as separate values;
- Present all type of rapid estimates as additional values of the time series vector of the target variable;
- Present all kind of rapid estimates as individual values;
- Do not have any clear policy for the presentation of rapid estimates.

Alternatives

(A) Present flash estimates as an additional value of the time series vector describing the target variable without revising any previous value; present nowcasting and advanced estimates as separate values;

(B) Clearly separate regular production of official statistics and rapid estimates by presenting all kind of rapid estimates as separate values;

(C) Always append rapid estimates to the time series of the target variable or do not have any clear policy concerning the presentation of rapid estimates.

19.8.3 Documentation of rapid estimates

Description

Rapid estimates have to be always accompanied by a complete documentation helping users in understanding their nature, the methodology used and their interpretation. Standard metadata structures associated to official statistics can be considered as the primary documentation tool for rapid estimates. Nevertheless, due to their nature it will be very relevant to produce also technical documentation on methodological aspects and also on empirical studies related to rapid estimates.

Options

- Incorporate documentation about flash estimates into the standard metadata linked to regular estimates of the target variable and produce specific metadata for nowcasting and advanced estimates;
- Use specific standard metadata for flash estimates, nowcasting and advanced estimates;
- Complement metadata files with methodological and empirical documentation on rapid estimates;
- Develop specific metadata files for rapid estimates not in line with the internationally agreed standards;
- Do not provide any documentation;

Alternatives

(A) Use the standard metadata files linked to regular estimates to document flash estimates too; use specific metadata files, following international agreed standards, to document nowcasting and advanced estimates; complement metadata information by technical documentations such as technical guides, papers and studies;

(B) Document flash estimates, nowcasting and advanced estimates using specific metadata files, totally or partially in line with international agreed standards, possibly complementing them with more technical papers and studies;

(C) Use non-standard metadata files to document rapid estimates or do not document them at all.

19.8.4 Uncertainty and rapid estimates

Description

Rapid estimates are by definition characterised by a higher degree of uncertainty than regular ones. This is particularly true for nowcasting and advanced estimates which are produced very early and with several gaps in the information set. The uncertainty of rapid estimates is, among others factors, determined by the uncertainty in input data, the uncertainty of the statistical or econometric model used and the uncertainty related to the estimation of missing data.

Analysing in a separate way various sources of uncertainty could be of high interest but it is a quite complex operation which cannot be carried out at each rapid estimate release. Looking at total uncertainty is easier and could be easily performed on regular basis.

Communicating about uncertainty, especially in certain economic phases, could be as important as communicating point estimates delivered by rapid estimates. This comes from being the level of uncertainty not constant over the time but tending to vary according to economic or political phases.

Ways of communicating uncertainty can vary a lot from purely descriptive presentations, such as ranges associated to each estimate, to more complex ones such as the dispersion measures or confidence intervals associated to rapid estimates. The most complete way remains nevertheless the use of density forecasts which return for each rapid estimate an ex-post empirical probability distribution describing the variability of the rapid estimate in a very complete way.

Options

- Produce density probabilities associated to rapid estimates;
- Associate to rapid estimates confidence intervals or other measures of dispersion;
- Associate to rapid estimates ranges of variation;
- Do not provide any measure of uncertainty.

Alternatives

(A) Produce and disseminate densities functions associated to rapid estimates together with a clear documentation on how to interpret them; use advanced graphical tools to disseminate densities such as fan charts;

(B) Associate rapid estimates with either confidence intervals or ranges of variation together with a clear documentation on how to interpret them;

(C) Do not provide any measure to evaluate the uncertainty associated to rapid estimate.

Bibliography

Barcellan R., Mazzi G.L., Montana G. and Ruggeri Cannata R. (2009). Glossary of statistical terms associated to the estimation of key economic indicators, International Seminar on Early Warning and Business Cycle Indicators, Scheveningen, The Netherlands. December 2009.

European Statistics Code of Practice (2011) revised edition, available on-line at: http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF

United Nations template for a Generic National Quality Assurance Framework (2012), available at https://unstats.un.org/unsd/dnss/QualityNQAF/nqaf.aspx

Getting in touch with the EU

In person
All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: http://europa.eu/contact

On the phone or by e-mail
Europe Direct is a service that answers your questions about the European Union. You can contact this service
– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
– at the following standard number: +32 22999696 or
– by electronic mail via: http://europa.eu/contact

Finding information about the EU

Online
Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU Publications
You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU law and related documents
For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU
The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

KS-GQ-17-008-EN-C

Handbook on Rapid Estimates

In an ever more interconnected and interdependent world, where time is increasingly of the essence, more reliable and up-to-date data are needed to enable and support the process of effective decision making, e.g. via rapid estimates. This handbook provides an overview of the different types of rapid estimates and the statistical and econometric methods for their construction. It also offers guidance for the implementation of rapid estimates in practice and their quality assessment.

For more information

http://ec.europa.eu/eurostat/

Publications Office

ISBN 978-92-79-68896-6