UNITED
NATIONS

E

Economic and Social Council

Distr.
GENERAL

CES/2001/21/Add.1
4 April 2001

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Forty-ninth plenary session
(Geneva, 11-13 June 2001)

Experience of the Statistical Office of the Slovak Republic with DOSME Project

Supporting paper submitted by the Statistical Office of the Slovak Republic

I.   Introduction

1.   The report summarises experience gained in the first phase with panel projects of newly created enterprises within years 1995 - 2000. It presents a general overview of the project, highlights some issues related to area sampling used in the project and finally it points out experience gained from the central data processing.

II.   Panel Projects of Newly Created Enterprises - an overview

2.   There were two consecutive projects on the demography of enterprises carried out in the period of 1995 - 2000, called PECO and DOSME. PECO is an acronym from French "Pays Europeannes Centrales et Orientales" and DOSME stands for "Demography of Small and Medium-sized Enterprises in Central European Countries".

3.   The following countries participated in DOSME project from the beginning: Albania, Bulgaria, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Rumania, Slovakia and Slovenia. In 1998, the Former Yugoslav Republic of Macedonia (FYROM) joined the project.

4.   Both phases of the project were co-ordinated by Eurostat and the experts from the statistical offices of the EU countries (France, UK, Netherlands, Finland and Denmark) assisted the project as consultants. The processing of the data collected by the countries participating in the project was provided centrally by the Institute of Informatics and Statistics - INFOSTAT (the research and development centre of the SO SR). INFOSTAT also maintains the full set of project documentation. The working group of the participating countries met twice a year to discuss project methodology and to

evaluate the progress. Several training sessions and other special working sessions were held at the beginning of the project.

5.   The main objective of the project was to observe the establishing and closing down of the enterprises, their attributes and activities, in order to obtain data on the size and characteristics of the active enterprise population and its development. As the panel had to be established on the basis of business registers of the national statistical institutes, analysis of the quality of registers in each country was included under the main aims of the project.

6.   One of the subordinate objectives of the project was to train the experts from the participating countries on methodology of sample surveys used during the whole panel survey within a large multi-national project. The transfer of knowledge and development of skills was concentrated in such a way that in future the CECs will be able to continue the panel of newly created enterprises independently and apply the methods to other areas of enterprise statistics. The knowledge and skills that have been developed include panel methodology and statistical methods such as sampling, imputation, estimation and grossing-up.

7.   Besides the identifiers, management and auxiliary data the central database, composed of data collected within the project by countries, contains three major groups of data:

- data from the register (sampling base, register at reference date)
- data as survey results
- imputed survey results

8.   Considering the differences between these groups of data and the basic structure of the questionnaire the following types of analyses could be carried out:

- structural analyses (numbers of units)
- analysis of quantitative data on enterprises
- profile of entrepreneurs
- problems in developing the enterprise
- quality of the register.

9.   Almost all these analyses are prepared regularly for our international publications and countries also provide some of them for their national publications.

III.   Area issues in the projects

10.   There were two types of the area related issues in the projects:
         area sampling
         area characteristics.

11.   In the area sampling a country could, for various reasons, prefer some small areas or regions. One reason may be to have well trained interviewers in a given region, another to define a region struck by a catastrophe. There may also be other specifics of the region that have a particular weight in a sample.

12.   The logic of area sampling and the use of area characteristics can be illustrated briefly by describing the involved variables.

13.   The area related variables in Project DOSME:

a) Initial region code (P05)

This code splits the regions of the country into two parts:
> 1 Region of the capital
> 0 Other regions

This code is stored in the sampling base and is not changed later. We call a region the largest of territorial division administratively defined within a country.

b) Initial zone code (P06)

This code divided the municipalities of the country into three groups:
> 1   Included in main administrative urban agglomeration of its region
> 2   Included in other urban agglomeration of its region
> 0   In rural area

This code is also stored in the sampling base and is not changed later.

c) Urban/Rural (PD6)

Urban/Rural code is derived from initial Zone code (P06) and distinguishes rural areas from urban clusters by the following condensed form:
> 0   Rural area
> 1   Urban clusters

d) Village inclusion (P07)

In case of area sampling it is a code indicating if the village is included in the sample with values:
> 1   included
> 0   not included

In the case of direct sampling the variable P07 has always value 1. The code is stored in the sampling base and is not changed later. This variable is used in computation of the variable P08 -"village weight".

e) Village weight (P08)

> In the case of area sampling not using the BERNOULLI method this variable indicates the weight of selected villages in relation to the global rural area.

> If the village is not included in the area sample (P07 = 0)   P08 = 0
> If the village is included (P07= 1), then P08 = N/n,   where N= number of enterprises in rural villages and n = number of enterprises in all selected villages.

> In the case of direct sampling P08 = 1,00. This number is stored in the sampling base and is not changed later.

f) Increasing factor (P09)

> There are two cases of increasing the sample size:
> In the first case of area sampling, part of the survey cost that is saved must be used to increase the size of the sample in rural area. This increasing factor must be the same for every sub-stratum in rural area. That is:

if **V-P06** = 0   than   **V-P09** = k
if **V-P06** = 1 or 2 than   **V-P09** = 1

The second increasing factor is valid for every case. Countries can decide to choose a higher sampling rate than the minimum sampling rate indicated by Eurostat. If they do so, the dominant factor must be determined for some sub-stratum defined by the criteria chosen by the country. For example, a country may decide to increase the sampling rate for enterprises with trade activity, others to increase the sampling rate in specific regions. If they don't want to increase the sample at their own cost, P09 = 1, this second dominating factor can be multiplied by the first one. This variable is stored in the sampling base and is not changed later.

IV.   Experience gained from the panel projects

14.   It is believed that the main aims of the project, i.e. a) to train the CECs countries on providing sample surveys and panel surveys and b) establishing an international longitudinal data base on the transition process, have been fully achieved.

15.   The main benefit of the project results from experiencing the whole duration of a panel project from designing a questionnaire through the field work to preparing publications from collected data files.

16.   Besides we can sum up the experience gained from the project in several fields.

a) In methodology of sample surveys and panel surveys:

- using a permanent random number in Bernoulli sampling
- the importance of sample allocation in a stratified sample survey
- using variables characterising the state of activity of units both in surveys and in the register
- using survey management data in data entry files

b) In planning and managing sample surveys and panel surveys in an international project:

- the existence of a steering committee or a co-ordination team is inevitable
- the necessity to involve into the project equally managers, statisticians, programmers and field workers
- consistent project documentation is necessary and preferably it should be kept on computer files

c) In managing the fieldwork of panel surveys

- in sample surveys mailing the questionnaires is not sufficient enough procedure, a personal contact in interview have to be also applied,
- involving interviewers also in indirect control of existence of units and identifying the state of activity of units,
- instead of a simple data entry, using a consistent entry package containing also record checks, file checks, table generations, etc.

d) In editing and imputing sample survey data

- there is a close relation between auto corrections and the imputation process,
- a review of various methods of imputing and choosing the suitable ones,
- experiencing the hot deck method of imputing missing data within a tailor made software package.

References to Eurostat publications:

Enterprises in Central and Eastern Europe, Office for Official Publications of the European Communities, 1996

New enterprises in Central and Eastern Europe in 1995 (Supplement 3, 1997, of the distributive trade and services series)

New enterprises in Central European Countries in 1996, Office for Official Publications of the European Communities, 1999

Enterprises in CEC: 1995 - 1997

New Enterprises in Central European Countries in 1997