



## 7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

### DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

**F. Data quality: Quality assurance measures**

**F2 – Comparison of LFS data with other sources**

Lenart Lah – Slovenia

## **The Comparison of the Labour Force Survey Data and those from the Administrative Sources in Slovenia**

The LFS in Slovenia has been facing many challenges in the last couple of years. On one hand, the data we produce have gained attention of policy makers and economic analysts, as Slovenia needs reforms on the labour market in order to battle the economic crisis and thus a wide range of data on labour market statistics. On the other hand, the available financial resources have been constantly decreasing, so we need to look for new, "less costly" ways to keep providing rapid, timely and reliable data. One of the most promising ways to expand the LFS data and at the same time to maintain the reliability of data on the same level is to acquire new methods of data collection, among them especially to introduce or elaborate the use of existing administrative sources.

Many questions arise, when we consider linking LFS data to other different sources, for instance:

- How accurate are LFS data compared to the data acquired from registered sources?
- Which data would it be reasonable to acquire from other resources?
- What are the pros and cons of partial acquisition of data from other sources?

Thus, the aim of the paper is to analyse the overlapping between LFS data and register data in order to get the answers to the stated questions. By making these comparisons, we will also get some feedback on the quality of the LFS data.

### **SOURCES**

In Slovenia, there are various sources of data on labour force and SORS (the Statistical Office of the Republic of Slovenia) is the biggest data supplier in this field. The data on labour force, used in the Labour Statistics Department at SORS, can be split into two groups according to the source of data:

- Household statistics:
  - Labour Force Survey
- Registers:
  - Statistical Register of Employment
  - Register of Unemployed Persons
  - Tax register
  - Central Population Register
  - Registered based census

#### **Statistical Register of Employment (SRE)**

SRE contains employment data on persons in employment on the individual level, such as school education, occupation, type of work and tasks, shift work, labour relation, working hours (weekly), etc. It also contains the personal identification number, which allows matching of these data on the individual level with those from other data sources.

In fact, SRE covers only persons in paid employment and self-employed persons with the employment contracts.

### **Register of Unemployed Persons**

This register is kept by the Employment Service of Slovenia.

Registered unemployed persons are persons aged 15 to 65 years whose health enables them to accept employment, are registered at the employment office and are willing to accept within 14 days employment which corresponds to their educational attainment, knowledge and capabilities. In addition, they are not retired, not imprisoned for more than 6 months, not already employed, not owners or co-owners of an enterprise with which they created income in the last calendar year and which they could use to support themselves or the income they created does not exceed the guaranteed earnings. Registered unemployed persons do not perform independent activity as a regular job, are not owners, not sharecroppers, not renters or other users of agricultural or forest areas or real estate which can provide a living.

### **Tax Register**

Data on wages are collected by tax administration, which is run by the Ministry of Finance. Data contain personal identification number so that we can match them to other data.

### **Central Population Register**

Individual data can be acquired from the Central Population Register (CPR) administered by the Ministry of the Interior. From 2008 on, the CPR has been integrated with the Register of Foreigners (until 2006 the Database of Foreigners), from which SORS initially acquired separate, and before 2003 aggregated, data on foreigners.

### **Census data**

In 2011, the first register-based census was carried out by SORS. Census data include data on many characteristics that can be directly linked to labour force topics. Recently, we have been discussing options of linking LFS to census data in order to acquire certain data (date of birth, education, country of origin, etc.) instead of interviewing respondents.

### **Methodological differences between the register data and the data obtained by statistical surveys of persons in employment and unemployed**

Register data and data obtained by statistical surveys differ regarding:

- **Source:** register data are based on the Statistical Register of Employment and the Register of Unemployed Persons while survey data are obtained by the Labour Force Survey. In case of register data there is full coverage, while Labour Force Survey results are estimates based on the statistical sample.
- **Reporting period:** register data are extracted on the last day of the month, while survey data refer to the activity of the respondent in the week before the interview (from Monday to Sunday).
- **Observation period:** register data are extracted on the last day of the month, while the Labour Force Survey is carried out quarterly.
- **Categories included in persons in employment:** register data cover only persons in employment with employment contracts, while the Labour Force Survey also covers unpaid family workers and persons working under contracts for work or for direct payment, i.e. persons who in the week (from Monday to Sunday) before the interview performed any work for payment (in money or in kind), for profit or for family gain.

- **Definitions of unemployed persons:** registered unemployed persons are persons who are registered by the employment office and fulfil all criteria defined by the employment office. According to the Labour Force Survey, unemployed persons are people who in the week (from Monday to Sunday) before the interview did not perform any work for payment (in money or in kind), profit or family gain but have been actively seeking work in the last four weeks before the interview and are prepared to accept it in two weeks. Unemployed persons are also those who found work and will start working shortly after the interview.

## THE PROCESS

As we can see, Slovenia has several administrative sources from where we can get not only the data already obtained by the LFS, but also the data not obtained by the LFS. For the purpose of this analysis, we will focus only on the analysis of data already obtained by the LFS. In the previous years, we linked LFS data to tax register data, in order to get information on employees' wages, which were not included in the survey. The experience and the results turned out to be very positive, so we decided to continue linking the LFS to other sources.

The process of linking data from various sources always starts with choosing the common identifier. In Slovenia every person has his/her own personal identification number (PIN), so the easiest and the most reliable way to link data is via PIN. Still, some sources, the LFS included, do not contain PIN data. Therefore, to make LFS data comparable, we firstly need to find and add PIN to each person in the survey. We use the process called matching; certain data from the LFS (first name, last name, date of birth, and gender) are matched with the same data from the CPR. Once the persons are found in the CPR, so is their PIN. In theory the process of getting PIN should be smooth and the results should be reliable, but some problems do arise. Some respondents do not want to disclose certain data (usually those of their birth), or intentionally or unintentionally do not disclose correct data (esp. instead of their real name they state their nickname), sometimes interviewers might misspell data, etc. These deviations aggravate the matching process as the data on some persons from the LFS do not perfectly match those from the CPR.

The main aim of this paper is to link the LFS data, data from the SRE and those from the Register of Unemployed Persons. Thus, we want to find out what status do persons from the LFS have in these two registers. Therefore we will find out the share of persons with the same status in the LFS and in register, while for those whose status does not fit either the survey or the register, we will seek for the reason which this is so. Besides the working status, these two registers include some other data that is also included in the LFS. We will compare these data too in order to test the overlapping and therefore the accuracy of the LFS data. As both, the SRE and the Register of Unemployed Persons, contain the PIN, matching with the LFS data is feasible.

## RESULTS

As we have already mentioned, the LFS data will be compared to the Statistical Register of Employment data and the Register of Unemployed Persons data.

According to the available data, we will compare:

- LFS with SRE by:
  - Working status of persons in employment
  - Working time per week
  - Educational attainment level
  - Full time/part time
  - Profession
  - Activity of work
  - Marital status
- LFS with Register of Unemployed Persons by:
  - Status of unemployed
  - Unemployment spell
  - Educational attainment level
  - Receiving unemployment benefit

The detailed comparison of the statuses and other indicators will be the subject of presentation at the Workshop in Madrid.