Distr.: General
17 April 2012

Original: English

Economic Commission for Europe

Conference of European Statisticians

UNECE-Eurostat Expert Group Meeting on Censuses Using Registers

Geneva, 22-23 May 2012
session 3: Availability, completeness and quality of data from registers and other sources

Use of Survey Data (LFS) to Evaluate the Quality of Register-based Census in Finland

Note by Statistics Finland[1]

I. Introduction

1. Statistics Finland has carried out a register-based census since 1990 and the production of census data is part of its annual statistics production. The use of the personal identification code is an essential part of all statistics production and gives a sound basis for quality measures.

2. Several approaches can be used to study the quality of registers. The Population Register Centre, the authority responsible for the Population Register in Finland, makes an annual survey concerning the quality of the address data in the Register. One important way of evaluating the quality of the data is to use Statistics Finland's household surveys, such as the Labour Force Survey.

3. The use of the personal identification code makes it possible to combine datasets at the individual level. This study focuses on differences between the census data and the Labour Force Survey data. It is useful to compare the employment data of the census and the Labour Force Survey because their definition of an employed person is the same. Most of the differences at the individual level are then more or less due to the different data collecting method, although some of them relate to a different reference period. Survey data and register data each have their own types of measurement errors. We have to recognise the weaknesses in both methods in order to be able to use all available data to improve the quality of statistics.

[1] Prepared by Jari Nieminen and Kaija Ruotsalainen.

.

II. Earlier studies of reliability of censuses

4. The first comparisons between register and questionnaire data were carried out with the 1980 census data. In 1985, an exhaustive comparison was made to determine how the census data on main type of activity and industrial status compared with the data extracted from registers. In addition, register data for 1987, 1988 and 1989 on employment have been compared with a corresponding sample survey. The results of all these comparisons indicated that the level of agreement between register-based data and questionnaire data was sufficiently high to justify the decision to carrying out the 1990 census entirely based on register data.

III. Evaluation of census 2010 data using the Labour Force Survey

5. Statistics Finland uses three different approaches when using survey data to evaluate the quality of register-based census data:

- Ad hoc survey question to control the quality of registers. An approach of this kind is the annual survey of the Population Register data.
- Analyses of the figures at aggregated level to get new information on changes in both collection methods and data sources.
- Compilation of data from register-based statistics and surveys at the individual level by using personal identification codes.

6. This paper gives three examples of how Statistics Finland uses the Labour Force Survey (LFS) data to evaluate the quality of the annual register-based production of census-related statistics. The data of the Labour Force Survey and the annual register-based employment statistics are matched by using the personal identification code. This compiled dataset has variables from both datasets. It is possible to make these comparisons and analyse reasons for differences in them. Both statistics are able to use this information in order to improve their process

IV. Address

7. The Population Register Centre is the only source of address data for persons in Statistics Finland's surveys. It is also crucial when producing the family and housing data in the census. Once a year, the Labour Force Survey contains an additional question concerning the quality and accuracy of the address data from the Population Register. This has been inquired annually since 1998. The proportion of people who report that their address in the Population Register is correct has been high right from the start. In 1998, over 96.8 per cent of the respondents had a correct address in the Population Register and in 2010 the share was 98.4 per cent. This figure is used as one of the key indicators on the Population Register Centre's scorecard, which means that they will make some efforts to maintain good quality also in future.

V. Type of family

8. There is seldom any need for imputations when registers are used in statistics production. In the register-based census we have to impute information for some variables, such as cohabiting. Cohabiting couples are identified by using information on sex, age and

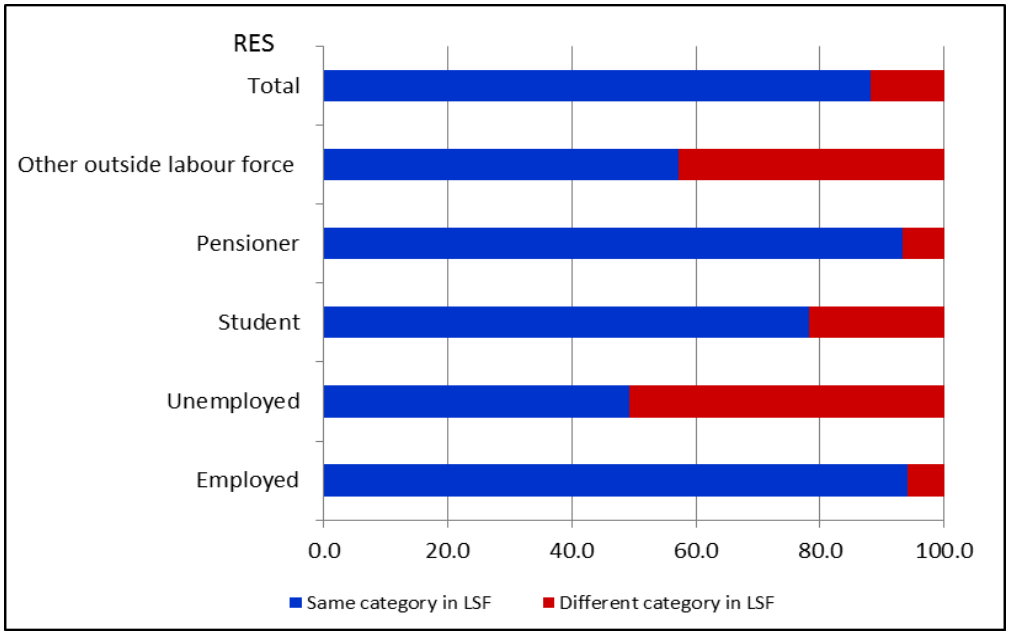
place of residence. This information is not available from any of the registers used in the census. To evaluate the quality of this imputation, we are able to compare the type of family in the census and in the Labour Force Survey. In family statistics, this method has been used since 1990. Labour Force Survey data from December 2010 were used in the comparison. According to the Labour Force Survey, 91.3 per cent of those who lived in a cohabiting partnership were classified in the same way also in the census. Of them, 9.7 per cent in the LFS were not in the same category in the census, 55 per cent lived alone and 35 per cent were married either with or without children.

VI. Main type of activity

9. The evaluation of the census 2010 data was carried out in 2012 when the last results on main type of activity were published from the census. These data will be analysed even further by using all key variables. The most complicated variable to build up in a register-based census is main type of activity. Almost 30 registers are used to update this variable in the census. We have used data for December 2010 to compare the Labour Force Survey with the census. The interviews of the Labour Force Survey were conducted in the five last weeks of 2010, near the reference point of time for census 2010, which was the last week of the year.

10. Comparisons between various years can be made because regional employment data are produced annually. When the years 2002 and 2010 are compared, there is a significant change in the number of persons classified in the same category. In 2002, 75.7 per cent were classified in the same category both in regional employment statistics (RES) and in the Labour Force Survey (LFS). In 2010, the percentage dropped to 67.6. Thus, the change is due to a risen rate of non-response in the LFS. If the non-response is not included in the comparison, the number of persons classified in the same category has fallen by only one per cent, from 89.2 per cent in 2002 to 88.3 per cent in 2010.

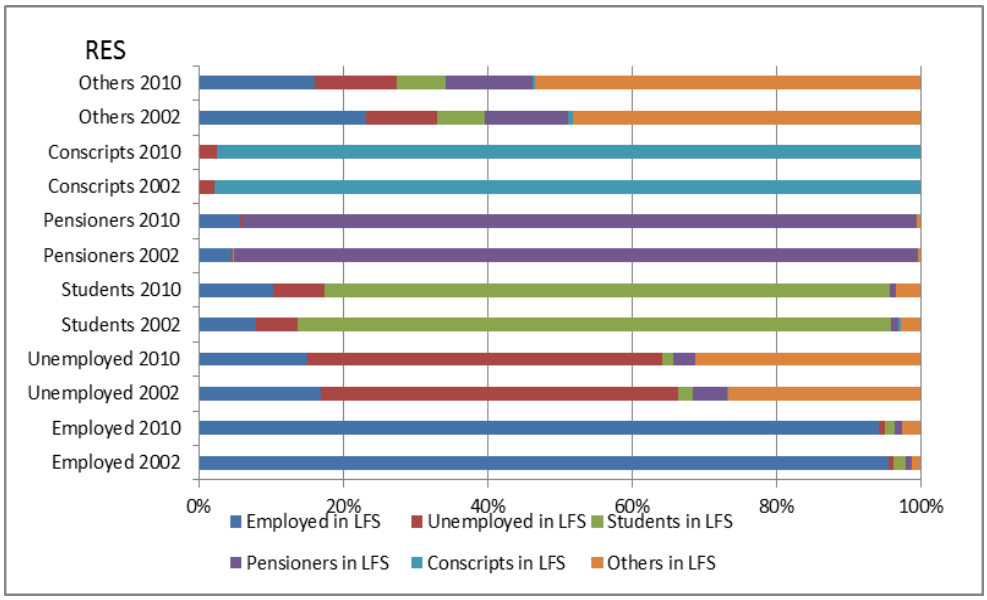
Figure 1: Main type of activity - Per cent classified in the same and different categories in the Labour Force Survey and the 2010 census (without non-response)



11. The proportion of the unemployed has constantly been higher in the RES than in the LFS. The difference is biggest in the category of unemployed. Only 50 per cent of those classified as unemployed in the RES were also classified unemployed in the LFS. In the RES, a person is classified as unemployed when he/she is registered as an unemployed job seeker in the register of the Ministry of Employment and the Economy. In the LFS, the activity is defined according to a respondent's own reporting.

12. It is important to evaluate these kinds of differences even further. In these data, we found that a job seeker who will be starting a job in the near future will sometimes be counted as employed in the LFS. This is because the new job will have actually begun after the reference week of the Survey. Almost 75 per cent of those who were unemployed in the RES but employed in the LFS had the end of the unemployment period within a month and 90 per cent within two months.

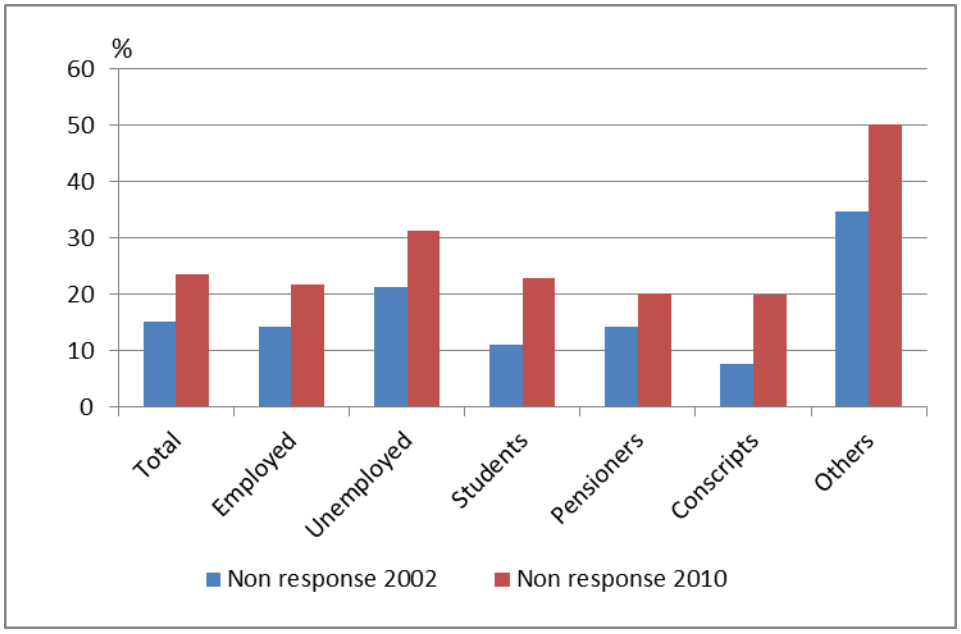
Figure 2. Percentages classified in various categories in regional employment statistics (RES) and the Labour Force Survey (LFS) in 2002 and 2010



13. Comparisons between the RES and the LFS can also provide important information on the structure of the non-response in the LFS. The non-response has been increasing in all categories of main type of activity. Group non-response was almost half in the category of others outside the labour force. The differences between the RES and the LFS have been at the same level since the annual production of regional employment statistics began in 1989.

14. It is possible to find out the reasons for most of the differences between the RES and the LFS by going through various registers.

Figure 3. The structure of non-response in the LSF (%) by main type of activity in regional employment statistics (RES) in 2002 and 2010



VII. Conclusions

15. The use of several registers and other administrative sources will reveal differences between them. Although the differences between the various sources are a challenge to the production process, comparisons between registers will provide important information on the quality of the registers.

16. Register-based census production and the annual production system of census statistics demand a new kind of approach to the quality work on census data. Instead of a traditional, post-enumeration survey there will be a need for continuous quality work throughout the whole production process. It is very important to monitor the quality of registers as a permanent part of data collection and data editing processes to minimise differences between register-based data and survey data.

17. In the register-based census, the use of sample surveys is an excellent way to replace a traditional post-enumeration survey.