The Regular Labour Force Survey as a Quality Survey of the Finnish 1985 Census

by Aarno Laihonen[1]
Central Statistical Office of Finland

The paper deals with experiences from using the sample-based regular monthly Labour Force Survey as a quality survey of census data on the economic activity of the population. The method uses record linkage at the micro level between the data of persons in the Labour Force Survey sample and the data of the same persons in the census file. An exact linkage is made possible by the uniform personal identifier used in the Finnish population registration system.

The use of the Labour Force Survey as a reference quality survey of the census was possible because the survey week and the census week coincided, and because the survey and the census measured the same variables according to the same concepts and classifications. In this way, savings were achieved in the cost of the census quality survey. For the purpose of the quality survey, after the regular survey interview a few additional questions were asked of a sub-sample consisting of one-fifth of the Labour Force Survey sample (about 2,300 persons).

To ensure as errorless results as possible, the data of the sub-sample were reprocessed after being entered and coded as usual. The additional information on the sub-sample was utilized when estimating the final errors on the basis of a micro comparison between the census data and the survey data of the whole Labour Force Survey sample.

Introduction

In Finland, modern population and housing censuses have been carried out in 1950, 1960, 1970 and 1980, and so-called mid-decade censuses in 1975 and 1985. The latest census of 1985 was carried out within extremely tight budget constraints imposed on the Central Statistical Office by the Ministry of Finance. The tight budget constraints also affected the production of census data. New cost-saving devices had to be used.

A general outline of the 1985 census

To give an idea of the cost frame, the direct costs of the 1980 census amounted to about 80 million marks (about 17 million US dollars) and those of the previous mid-decade census of 1975 to about 26 million marks, in 1986 prices. The total expenditure of the 1985 census was not to exceed 18 million marks.

The central point of departure for the planning of the system solutions of the November 1985 census was to minimize the amount of manual work in census data collection and processing.

First of all, data collection was minimized by an extensive use of registers and administrative records when gathering the basic census data. This was largely made possible by the comprehensive, high-quality population registration, taxation and social security systems characteristic of all the Nordic countries. The use of registers and administrative records in population and housing censuses has increased steadily since the 1970 census. This development has been aided by the widespread use of the uniform personal identifier in different registers and administrative records.

A significant improvement in the register situation, which helped a great deal in bringing down the cost of the latest census, was the establishment of a building and dwelling register for Finland on the basis of the 1980 population and housing census. This register is operated in connection with the Central Population Register, and it allows the linkage of persons and dwellings. Therefore, no questions on housing were needed on the 1985 census form.

The census of November 1985 used only one questionnaire, namely for gathering employment data. The Central Population Register was used as the mailing list for the population of working age. *The questions and instructions on the census form are presented in Annex 1.*

The questionnaires were sent out by mail from the Central Statistical Office (CSO) and were returned by mail to the CSO. No local census organization was used. Another special feature of this census was that the census form was preprinted, not only with the respondent's name and address but also, for about half the population, with the name of the respondent's workplace as it appeared in the 1980 census and with the respondent's occupational title obtained from the Central Population Register. Persons obligated to respond needed only to report any changes or errors that had occurred in this information.

The census forms went directly to data entry, which was carried out as key entry. Naturally, only changes and additions on the forms had to be keyed. In this way

complete 'pictures' of the census forms were converted to machine-readable form before any other processing operations were performed. This enabled batch mode checking and correcting of the form data, leaving only about 10 per cent of the forms to be checked and corrected manually, on terminals. Next, extensive automatic coding was applied to workplace and occupation data. Thus, the number of forms requiring manual processing was drastically reduced in all phases of processing. This reduced the total cost of the census and allowed preliminary publication of the most essential census data as early as December 1986.

In the final phase of data collection, register data were also used to obtain, by imputation, the census form data of non-respondents. In this way, a satisfactory 98.6 per cent total coverage was achieved for the central data of the census form. This also contributed to the relatively small regional variation in coverage, even though the census was carried out as a direct mail-out, mail-back system without any local census organization and with only one reminder sent to non-respondents. Register imputation of questionnaire data was tried for about 139,000 persons (3.7 per cent of the population of working age), 84,000 of whom were non-respondents and the rest persons whose responses were incomplete.

The regular Labour Force Survey, the study week of which coincided with the census week, and the 1985 household survey were used as quality surveys of the census. According to the quality surveys, the general quality of the 1985 census data is significantly better than that of the previous mid-decade census of 1975. The quality of employment data is in part slightly inferior to the quality of employment data in the 1980 census.

The setup of the quality surveys of the 1985 census

Because of the high-quality, up-to-date information obtainable from the Central Population Register (CPR) in Finland, the main purpose of the population and housing census is not to count the population, but to produce data on the economic activity and housing conditions of the whole population. The resident population of the country as registered in the CPR was the population of the census.

Thus, from the point of view of the quality of the census data, there was not, by definition, any undercount or overcount of the population. The aim of the census quality surveys was to analyze the quality of the data produced on different attributes of persons, dwellings and buildings. However the problem of under- and overcount was still relevant for dwellings and buildings because of the shortcomings of these data in the CPR.

The quality surveys of the 1985 census fall into two categories: those analyzing the quality of the CPR data (especially the data on household-dwelling units and dwellings) and those analyzing the census form data on the economic activity of the population. The data on household-dwellings units and dwellings, for instance, were analyzed by comparing the census data with corresponding data from the 1985 Household Budget Survey,

an interview-based sample survey of 12,000 households. Another source of dwelling data was a sample survey of dwellings registered as unoccupied in the CPR. This survey provided information on the overcount of dwellings in the CPR.

The quality of the data on the economic activity of the population was analyzed with the help of processing error studies (data entry errors and errors in coding and editing) and a special quality survey in which the final census data on persons were compared with the checked and corrected data of the interview-based regular Labour Force Survey. Some experiences from this survey and the methodology of the survey will be discussed in this paper. A short general description of the regular Labour Force Survey will be presented in the next chapter.

The Finnish monthly Labour Force Survey

The Finnish Labour Force Survey (LFS) is a sample survey based on a random sample of 12,000 persons selected from among the population aged 15-74 years. Data collection takes place mainly in personal interviews carried out by the CSO's interview organization.

The person interviewed is asked questions about his labour force participation (current activity), employment, unemployment, workplace, occupation, industrial status, time use, days and hours actually worked, overtime and secondary jobs, and normal hours of work. About 94 per cent of the interviews are telephone interviews and five per cent personal interviews. About one per cent of the answers are obtained using a mail questionnaire. The average non-response rate of the survey is about 4.7 per cent.

Structurally, the survey is a so-called continuous panel survey. The monthly sample of 12,000 persons is broken down into five rotation groups, each of which contains 2,400 persons. In the monthly survey, each rotation group can be considered to be an independent random sub-sample of the whole monthly sample. Each rotation group is surveyed five times over a period of 15 months. The lag between interviews is three months, with the exception of the lag between the third and the fourth interview which is six months. For estimation purposes, the sample is post-stratified by geographical area, age group and sex. The reference period of the survey is one week.

The Labour Force Survey as a frame of the quality survey

The monthly LFS made an excellent frame for the quality survey of the census form data for the following reasons:

1. As pointed out above, the census and the LFS used the same concepts and definitions in measuring the current activity of the population.

2. The study week of the LFS coincided with the census week.

3. The high response rate of the LFS, combined with more thorough interview questions and manual process-

ing by a more highly qualified personnel, made it reasonable to assume that, on the average, the final LFS data on persons were closer to the "true" values than the corresponding census data.

4. The uniform personal identifier used in Finland allowed easy and errorless linkage of the census records with the corresponding LFS records.

One important reason for using the LFS was that it was cheaper than a special separate quality survey like the one carried out in connection with the 1980 census.

The principal variables of the census selected as the topics of the quality survey were as follows:

1. Current activity of the population

    subdivided into

        Labour force

           Employed

           Unemployed

        Persons not in the labour force

           Conscript members of the defence forces

           Students and pupils

           Pensioners

           Home-makers

           Others

2. Industrial status

    subdivided into

        Wage and salary earners

        Own-account workers

        Employers

        Unpaid family workers

3. Employer sector

    subdivided into

        Private sector

        Central government (incl. provincial gov ernment)

        Local government)

4. Industry

    A national version of the ISIC.

5. Occupation

    A national version of the ISCO.

Because of the quality survey, it was necessary to take a few extra measures to check and to better the LFS data in order to make the "true" values as good as possible. In order to further reduce the cost of the quality survey, these extra measures were restricted to the first LFS rotation group, i.e., the sub-sample of 2,344 persons not interviewed earlier in the LFS.

The extra measures were as follows:

1. Additional questions (a total of 70, in addition to the standard 25 LFS questions), specially formulated from the viewpoint of the quality survey, were asked by the interviewers in connection with the LFS data collection. The collection of the data of the first rotation group was delayed by a few days in order to allow the group to respond to the census questionnaire before the LFS interview. The non-response rate for the additional questions was 10.2 per cent, compared with 5.4 per cent for the regular LFS questions asked of the first rotation group.

2. After the usual data processing operations, the data of the first rotation group were rechecked and recoded by persons not involved in the original checking and coding of the questionnaires of the first rotation group. The data from the additional questions were utilized in this process.

The results of the rechecking and recoding demonstrated the overall high quality of the LFS data. Concerning current activity, only 0.7 per cent of the original LFS values had to be corrected. The situation was similar with respect to industrial status. When it comes to the industrial and occupational classifications, the percentage of erroneous codes was somewhat higher, 4.3 and 4.0 per cent, respectively, of the number of employed persons.

Estimation of "error" in the census data

The usual procedure in this kind of quality survey setup is to match the quality survey records with the census records of the corresponding persons and to calculate the gross and net errors of the study variables as follows:

Gross error of a variable =

    100 per cent minus the percentage of similarity classified cases

Net error of a variable =

    the relative difference between the marginal distributions of the variable in the census and those in the quality survey.

This procedure could be applied directly to the sample of the first LFS rotation group. In this case, however, there were even uncorrected data on the whole LFS sample which correlated highly with the corrected "true" values of the first rotation group. Therefore, error estimates could be compiled using the so-called difference estimator (a special type of linear regression estimator see e.g. Madow, W.G.: "On some aspects of response error measurement", Proc. of Social Stat. Sc., American Statistical Association, 1965, pp. 182-192).

The difference estimator is of the form

$$Y = y + (x' - x)$$

where

x = the percentage of cases where the census values are identical with the original values of the first rotation group

y = the percentage of cases where the census values are identical with the checked and corrected values of the first rotation group

x'= the percentage of cases where the census values are identical with the original values of the whole LFS sample

The estimate of the gross error of a variable is given by $1 - Y$ and the variance of the estimator is of the form

$$V(Y) = y$$

$$(1/n - 1/N)S_y^2 \qquad +$$

$$(1/n - 1/n') (S_x^2 \qquad - 2S_{xy}$$

where

n = the sample size of the first rotation group

n' = the sample size of the whole LFS

S_x^2 = the population variance of variable x

S_y^2 = the population variance of variable y

S_{xy} = the population covariance of variables x and y.

In this way, the average of the standard errors of estimates could be reduced by about 50 per cent as compared with the standard errors for the sample of the first rotation group only. To have reached the same level of standard error by ordinary estimation methods, at least four times as large a sample should have been asked the additional interview questions and, consequently, at least four times as many data should have been rechecked and recoded.

[1]Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15-18, 1989.