INSTITUTO NACIONAL DE ESTADISTICA

Use of the Administrative Files in the Living Conditions Survey

Madrid, May 2014

Index

1. Introduction

The Living Conditions Survey (LCS) is an annual household survey conducted by the INE, and whose main objective is the systematic production of statistics regarding household income and living conditions. This survey, which is harmonised within the European scope via Community Regulation, enables ascertaining the level and the composition of poverty and social exclusion.

One of the most important difficulties that arise in surveys targeting households that are conducted by personal interview is the obtaining of information relating to household income. This type of variable traditionally presents higher non-response rates, and in many cases, it is necessary to resort to imputation procedures in order to be able to reconstruct total household income.

Access to administrative registers entails a good opportunity to improve the quality of the corresponding data, and in the efficiency of the collection method. The link to the microdata level, of those persons belonging to the sample, and that available data in the Tax Sources or in Social Security, enables ascertaining most of the income components in a detailed way.

In this sense, it is necessary to consider the sectorial recommendations of the High Council on Statistics that are found in the document of Proposals and recommendations for the formulation of the 2013-2016 National Statistical Plan. All of this is intended to improve the efficiency of the statistical production processes and to reduce the workload that it imposes on respondents. The current agreements with the State Tax Administration Agency, Social Security, the Tax Treasury of Navarra and the Delegation of Bizkaia in statistical issues are allowing for the exchange of information. In the case of the Living Conditions Survey, this has already been carried out for several years for performing comparative studies.

Likewise, within the European Union area, the use of administrative data in statistical production is an essential element in the modernisation process of social statistics. The new community legislation tries to facilitate the statistical use of administrative data.

Worth noting is that numerous countries use administrative files in this survey for the production of the income variables. In addition to the Nordic countries, the Netherlands, France, Austria and Slovenia make considerable use of the administrative files. In other cases, partial use is made, mainly covering social benefits.

2. Objective

The objective is to improve the production of the variables relating to household income, and to reduce the response burden, via the use of administrative files.

From the beginning of the survey in 2004, the collection method has mainly been the personal interview of the members of each of the target households of study. The interviewer visits the households and requests the information necessary to complete the questionnaires, making the pertinent number of visits to each household to collect all of the information required. In this way, the information is also collected regarding the income, as well as the rest of the variables relating to the living conditions of the household.

With the methodological change, the goal is to collect the income variables, mainly from the administrative files. Starting with the tax identification number (NIF) of those persons belonging to the sample, data is collected from the Tax Sources and from Social Security, which together with the data collected from the questionnaires, will enable constructing the income variables.

3. The global strategy for methodological change

The strategy consists of crossing the data from the survey with the data available in the Tax Sources or in the Secretariat of State for Social Security, to construct most of the components of income.

In the construction of the income variables by component, using the administrative files, in a first approach, the methodology would consist of identifying the corresponding cell or cells, applying the appropriate treatments, and thus obtaining the final variable required.

In some cases, this procedure is valid, but in other income components, it is observed that it is difficult to achieve an integral and sole use of the administrative files in their preparation. This is due to the fact that the problems presented, with regard to geographical coverage (problems in accessing tax data), coverage of income recipients (problems in certain groups, such as domestic staff, informal economy, etc.) or with the exact identification of the type of income in the administrative file.

A mixed methodology is applied, in which, one part of the information is from administrative files, and the other part comes from the questionnaire. Therefore, in some cases, it is necessary to maintain the questions from the questionnaire that gather the household income.

As the availability of data in the administrative files depends on compliance with certain conditions, some selective filters will be introduced into the questionnaire, which will cause the flow to omit certain sections when the respondent meets come requirements. These filters are expected to be implemented in the 2015 survey, which will imply a substantial reduction in the response burden.

Once the survey data is collected, which takes place between March and June each year, it is integrated with the administrative data. To this end, a filtering is performed of the NIF collection, which is submitted to the corresponding body that manages the administrative file.

Finally, combining the data from the survey and the administrative files, we proceed to construct the income-related variables. In a few cases in which it is not possible to collect the value from any source, it is necessary to use mathematical imputation.

4. Scope

POPULATION SCOPE

The target population object of study for the survey are those persons who are members of private households who live in main family dwellings, as well as said households. Those persons aged 16 years old and over on 31 December of the year prior to that of the interview are researched comprehensively using an individual questionnaire.

The collection of administrative data is performed through the NIF. The collection of the NIF in the LCS was introduced into the 2009 survey, with said variable being available for more than 98% of adults.

TIME FRAME

The variables relating to income refer to the calendar year prior to that of conducting the survey. This is called the income reference period. Therefore, in the survey conducted in T, normally between March and June, the income received from January through December in T-1 is recorded.

GEOGRAPHICAL SCOPE

The geographical scope of the survey is made up of the whole of Spain. With regard to using the administrative registers, in the case of given income components that are obtained from tax sources, it is possible that some territory will be excluded (access to tax data in Álava and Guipuzkoa). In these cases, the information will be collected via personal interview.

5. Impact of the methodological change

To date, basic comparison studies have been carried out in the INE between the data from the LCS and data from administrative files. Worth nothing is the Working Document available on the INE website, "Linking data from administrative records and the Living Conditions Survey"[1], in which a basic descriptive comparative analysis is performed, as well as assessing the impact that the use of administrative files in some LCS indicators would have.

The main conclusion of this study carried out with data from the 2007 survey is that the use of administrative files does not seem to have a significant impact on the indicators based on income distribution. However, it does have an impact on indicators based on income level, increasing its value significantly.

Other preliminary studies carried out with data from the surveys for 2009, 2010, 2011 and 2012 (the income refers to the year prior to the year of the survey) have yielded the following:

LCS indicators. Comparison of the survey and administrative files

	2009		2010		2011		2012	
	Survey	Admin. Files	Survey	Admin. Files	Survey	Admin. Files	Survey	Admin. Files
At-risk-of-poverty or social exclusion rate	24.5	24.7	26.7	26.1	27.7	26.7	28.2	27.2
Under 16 years of age	29.8	31.9	32.1	32.6	32.3	31.6	32.8	31.4
16 to 64 years old	23.3	23.0	26.7	25.3	28.2	27.0	30.1	29.0
65 years old and over	24.3	24.9	21.4	22.9	20.9	21.2	16.6	16.5
At-risk-of-poverty rate	20.1	20.4	21.4	20.7	22.2	20.6	22.2	20.8
Under 16 years of age	26.5	28.9	28.3	28.8	28.7	27.2	28.9	26.9
16 to 64 years old	17.9	17.5	20.1	18.6	21.3	19.3	22.4	20.9
65 years old and over	23.1	23.8	20.5	21.8	19.5	19.8	14.8	14.8
Gini Coefficient	33.0	32.9	34.4	33.5	34.5	34.0	35.0	34.2
S80S20	6.4	5.9	7.2	6.2	7.1	6.3	7.2	6.5
Average income per consumption unit	14483	17042	14369	16922	13907	16280	13885	16119

In analysing the indicators for the at-risk-of-poverty rate, it must be borne in mind that the confidence intervals are quite broad (Eurostat initially published this type of indicator without decimals). For example, in the case of the total for the population, the interval is approximately ±1.3, with a confidence level of 95%.

It is observed that, in recent years, the two sources have shown an increase in the value of the inequality and poverty risk indicators. Despite the trends coinciding with both sources, the increase is lesser in the case of administrative sources.

[1] Document available in www.ine.es/en/ in the Working Documents section

With regard to the income level, it is observed that the administrative source gives a significantly higher value.

6. Analysis of the administrative files

6.1. Registry of Public Social Benefits (RPSB). Social Security

Social Security manages numerous social benefits, with abundant information that also covers the entire national territory (in this case, there is no exception for the tax sources with regard to the "foral" regime Autonomous Communities). However, there are some benefits, mostly non-contributory pensions, that are managed by the Autonomous Communities, and that would remain outside the management scope of Social Security.

Nonetheless, Social Security is responsible for the management and functioning of a Registry of Public Social Benefits[2]. The creation of this register allows for coordinated knowledge and data provision among the affected entities and bodies, for the purpose of providing recognition of the benefits, as well as controlling and maintaining the right thereof and avoiding fraud.

Therefore, to begin with, through this register, Social Security is the depository for data relating to the benefits paid, both by managing entities of the Social Security System, and by managing entities of benefits from outside of the Social Security System.

In the case of the social benefits, a very precise statistical classification must be followed. In the case of the LCS, a classification based on the SEEPROS system (European system of integrated social protection statistics) must be followed, which harmonises the presentation of the social protection data.

6.2. Models 190 and 100. Tax Sources

To begin with, the information contained in the Income Tax Returns is quite rich, in terms of ascertaining the different components of the income of the households in the sample. However, initially, some difficulties may arise. In particular, the group of those persons not obligated to submit a tax return is very important, and on the

[2] More information from this register may be found at
http://www.seg-social.es/Internet_1/Normativa/index.htm?C1=1001&C2=2010&C3=3037&C4=4023

other hand, the possibility of a joint tax return may hinder the individualisation of income that the LCS requires.

For this reason, it is fundamental to access other information available in the Tax Sources. It must be considered that the liquidation of income tax is actually an annual adjustment of the taxes that have already been paid via a withholding and prepayment system.

Aside from the Income Tax Returns, the Tax Sources have extensive information. Specifically, they have "the statement of the annual summary of withholdings and prepayments from the Income Tax" (model 190), which contains individual information regarding different types of income, and which includes recipients of income who are not obligated to file a tax return. In the regulations referring to the model, there is detailed information regarding said model[3].

This register contains useful information for the construction of the income of the person working for others, and given social benefits. The data available for said register contains some fields that identify the income component (wages, unemployment, benefits, etc.).

Model 100 contains the Income Tax Returns. Though in principle, model 190 will be used as the main source of information, in the case of income of persons working for others, capital and property income, the information available in the return is used [4].

[3] More information from model 190 may be found at http://www.agenciatributaria.es/AEAT.internet/Modelos_formularios/modelo_190.shtml

[4] More information from model 100 may be found at http://www.agenciatributaria.es/AEAT.internet/Modelos_formularios/modelo_100.shtml