



Working Papers

01/2011

**Linking data from administrative records and
the Living Conditions Survey**

José María Méndez Martín

Pilar Vega Vicente

The views expressed in this working paper are those of the authors and do not necessarily reflect the views of the Instituto Nacional de Estadística of Spain

First draft: March 2011

This draft: March 2011

Linking data from administrative records and the Living Conditions Survey

Abstract

The Encuesta de Condiciones de Vida (Living Conditions Survey, LCS) is an annual survey compiled by the Instituto Nacional de Estadística (Spanish National Statistics Institute, INE). Access to administrative records offers a good opportunity to improve the quality of the relevant data and allow the use of a more efficient collection method. This paper offers a comparative analysis of different income components by linking the survey data with available data from the Spanish Tax Agency or Social Security system.

Keywords

Living Conditions Survey, administrative records, household income

Authors and Affiliations

José María Méndez Martín and Pilar Vega Vicente

Directorate of Sociodemographic Statistics, Spanish National Statistics Institute

Linking data from administrative records and the Living Conditions Survey

José María Méndez Martín jmmendez@ine.es, Department Head of the Living Conditions Survey at the Instituto Nacional de Estadística

Pilar Vega Vicente pilar.vega.vicente@ine.es, Programme Director at the Instituto Nacional de Estadística

Abstract

The *Encuesta de Condiciones de Vida* (Living Conditions Survey, LCS) is an **annual survey** compiled by the Instituto Nacional de Estadística (Spanish National Statistics Institute, INE). The primary aim of this survey is the systematic production of statistics on **household income** and living conditions. The survey, which is harmonised at European level through a Community Regulation, allows us to find out the level and composition of **poverty** and social exclusion.

Access to **administrative records** offers a good opportunity to improve the quality of the relevant data and allow the use of a more efficient collection method. This paper offers a **comparative analysis** of different **income components** by linking the survey data – at microdata level using the Spanish Tax ID number (NIF) – with available data from the Spanish Tax Agency or Social Security system.

Keywords: Living Conditions Survey, administrative records, household income.

1. Introduction

One of the major stumbling blocks with household surveys that involve personal interviews is that of **obtaining information on household income**. Traditionally, this type of variable has one of the highest rates of non-response and it is often necessary to use imputation procedures to reconstruct total household income.

Access to administrative records offers a good opportunity to **improve the quality** of the relevant data and allow the use of a more **efficient collection method**. The link between the individuals in the sample and the data available at the **Tax Agency** or the Department of **Social Security**, at microdata level, would provide us with detailed information on the majority of income components.

There are several methodological issues that need to be addressed when accessing this type of data, including the **availability of a NIF** (common variable of identification in the LCS and the administrative records) and the mapping of the concepts used in the LCS on to those of the administrative sources.

Until the 2008 LCS, the data collection process did not include the entry of NIFs. A list of households was used for data collection, to which a reference person was assigned. For this study, data from the **2007 LCS** were used and the **NIF was assigned afterwards**. It was possible to obtain NIFs in approximately 80% of cases. These records were linked with Social Security data on social benefits and with data from the Tax Agency on different income components.

Since the 2009 LCS, data collection has been adapted to make use of the municipal register of inhabitants, indicating the people registered in the household (with their associated details, full name, date of birth, NIF, etc). A NIF will be available for approximately 98% of adults.

This study makes a **comparative microdata analysis** of a selection of household income components using data from the 2007 LCS. The information collected in the survey is compared to the data available in the administrative records. A study is attached at the end on the impact of the use of administrative records on the basic indicators obtained from the LCS. The results presented here should be interpreted with caution due to their partial coverage, given NIF availability in the 2007 survey.

We would like to thank the Spanish Tax Agency and the Department of Social Security for their invaluable assistance in providing the necessary information for this study. We would also like to express our gratitude to the various units of the INE for their support in this project.

2. Analysis of Social Security information

2.1. Information from the Social Security system

The Social Security system handles numerous social benefit payments and has access to a wealth of information covering the entire country (in this case, there are no exceptions as there are with tax sources with regard to Autonomous Communities with the charter system). Nonetheless, a number of benefits, mainly non-contributory pensions, are managed by the Autonomous Communities and are thus outside the scope of the management of the Social Security system.

Notwithstanding the above, the Social Security system has the task of managing and running a **Register of Social Security Benefits**, and it is this register that provides data on the benefits paid both by management agencies of the Social Security system and benefit management agencies that do not form part of the Social Security system.

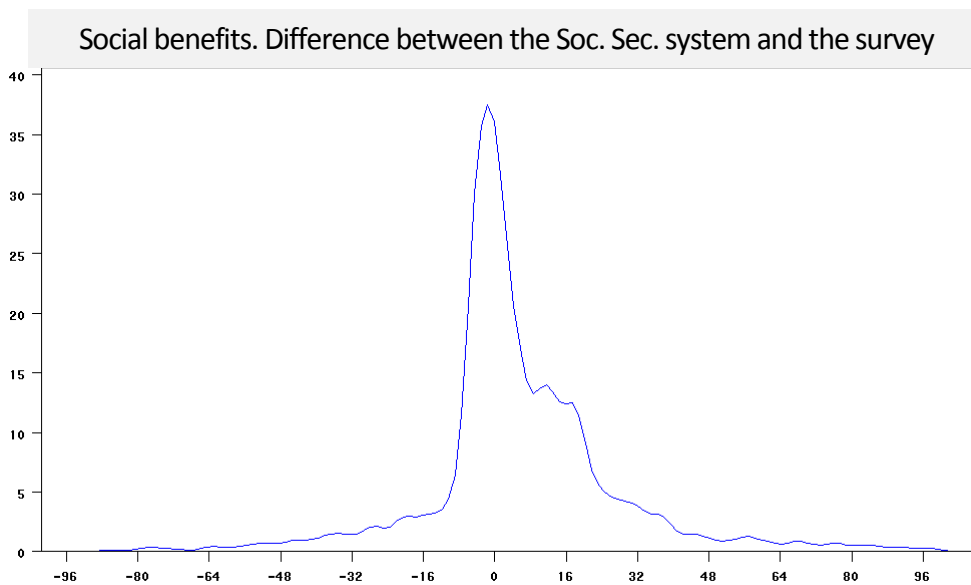
As with the tax sources, a very precise statistical classification must be adopted for social benefits. In this case, the social benefits included in the LCS must be converted following a classification based on ESSPROS (European system of integrated social protection statistics), which harmonises the presentation of data on social protection.

2.2. Comparative analysis

In the case of the Social Security system, information from the 2007 LCS was linked with Social Security data on the social benefits paid to people aged 65 and over (NIFs were available for 82% of this group).

In the first analysis, differences are observed in the type of benefit received. For example, some benefits are considered by the survey to be non-contributory old-age benefits, while Social Security records consider them to be contributory old-age or survival benefits.

Comparison of amounts. **We can see a certain underreporting in the amounts of social benefits included in the LCS**, as shown in the graph of the distribution of the relative difference, at microdata level, between the value of the amount in the administrative file and the value of the amount in the survey.



3. Analysis of Tax Agency information

3. 1. Information from the Spanish Tax Agency

In theory, the information contained in personal income tax returns is detailed enough to work out the various components of income for the households in the sample. However, there may be some difficulties: firstly, there is a rather large group of people who are not required to file returns and, secondly, the possibility of filing joint returns can make it difficult to identify individual incomes, which is almost always necessary with the LCS.

As a result, we require access to other information available at the Tax Agency. Besides personal income tax returns, the Tax Agency has a series of self-assessment forms containing very valuable data and information models presented by withholders, which even include tax-exempt income or income on which no withholdings have been made.

Specifically, the information supplied to INE in this study includes:

- **Filed returns (individual and joint).** These returns contain data on income broken down into different components.
- **Imputed individual returns** (or individual tax information). These contain individual information for certain sources of income, based on information from the Tax Agency.

The geographical scope is Spain with the exception of the Basque Country and the region of Navarre.

3.2. Comparative analysis

3.2.1 Interest, dividends and profits from capital investment

All adults were taken from the survey (28,656). By eliminating those residing in Autonomous Communities with the charter system (this leaves 26,237), this gave us a coverage of NIF availability of 79%, or 20,677 people.

Investment income is analysed on a per-household basis. Hence, we selected the households in which a NIF was available for all of its adult members. This gave a total of 15,804 people (76% of the previous figure).

If we exclude small amounts, we see that **a large percentage of households claiming to have no investment income in the survey actually do according to the Tax Agency.**

Some households also indicate in the survey that they have income from investments, but actually do not according to the Tax Agency. This is possibly due to the inclusion of investment funds, which the Tax Agency considers as capital gains.

Table 1: Distribution of households by investment income (LCS and Tax Agency) (income over EUR 100) (sample data). Horizontal percentages

	Tax Agency			
	Number of observations	Total	T1. With investment income	T2. Without investment income
Survey				
E1. With investment income	1,052	100.0	83.7	16.3
E2. Without investment income	6,271	100.0	34.7	65.3
Total	7,323	100.0	41.7	58.3

Comparison of amounts. If we analyse the distributions of the two sources, we can see a **significant underreporting in the amounts of investment income in the survey.**

3.2.2 Employee income and self-employment income

In this analysis, to avoid any overlap with social benefits (which also have the consideration of earnings from employment in personal income tax returns), we selected from the survey all people aged 18 to 64 years who stated that they were employed or self-employed for all 12 months of the year and did not receive social benefits. This gave a total of 12,047 people.

Of this figure, those residing in Autonomous Communities with the charter system were eliminated (leaving a total of 10,954 individuals). This gave a coverage of NIF availability of 79%, which left 8,613 people in the end. The analysis in this section is on a per-person basis.

Earnings from employment can be classified as earnings from salaried employment (employee income) or as earnings from self-employment (self-employment income). There is not a complete correspondence between the two sources for this classification, since some businessmen and women in the LCS set up companies but are listed as employees by the Tax Agency. It is also possible that workers who are self-employed according to the Tax Agency and who, for example, work for a single client, may be seen as salaried employees in the LCS.

Table 2: Distribution of individuals by earnings from salaried employment or self-employment (LCS and Tax Agency) (sample data). Percentages

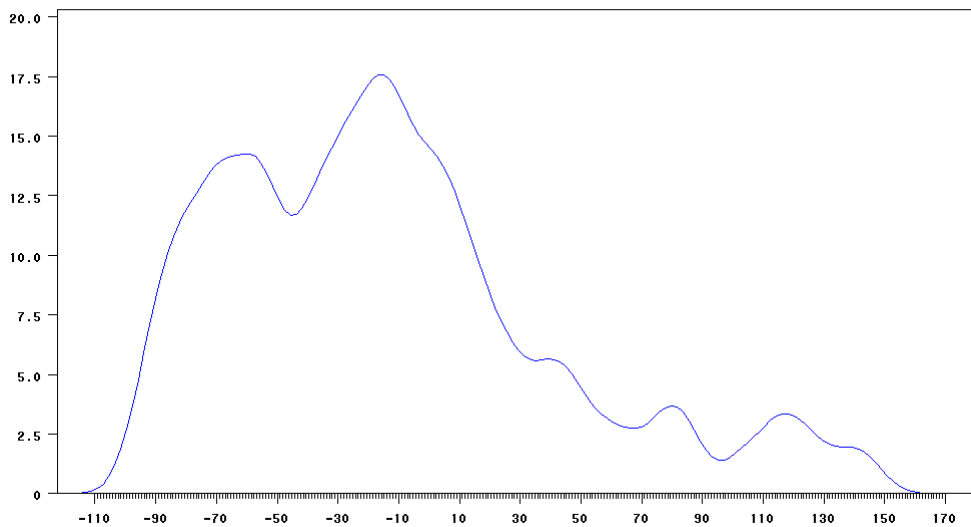
	Tax agency				
	Total	T1. Only earnings from salaried employment	T2. Only earnings from self-employment	T3. With earnings from salaried employment and self-employment	T4. No earnings from work
Survey					
E1. Only earnings from salaried employment	80.9	71.2	0.6	4.7	4.4
E2. Only earnings from self-employment	14.9	3.2	8.0	2.3	1.4
E3. With earnings from salaried employment and self-employment	1.5	0.4	0.2	0.8	0.0
E4. No earnings from work	2.7	0.5	1.3	0.4	0.5
Total	100.00	75.3	10.1	8.3	6.3

A separate comparative study will now follow of earnings from salaried employment and self-employment.

Self-employment income

Comparison of amounts. We can see a **significant underreporting in the amounts of earnings from self-employment of the Tax Agency**, as shown in the graph of the distribution of the relative difference, at microdata level, between the value of the amount in the administrative file and the value of the amount in the survey.

Self-employed. Difference between the Tax Agency and the survey



Note that in the case of objective tax assessment (modules system), the “net reduced earnings” were taken as income for the Tax Agency, although these are actually an imputation of profit from the activity.

Employee income

For earnings from salaried employment, a **separate study of the formal and informal economies is conducted**,¹ given that a different behaviour is detected. In the case of the formal economy, regular earnings lead to a similar situation to that of social benefits. In the case of the informal economy, the situation could go in the direction of earnings from self-employment.

Comparison of amounts. An **underreporting is seen in the salary amounts of the Survey in the formal economy** and a **slight underreporting is seen in the salary amounts of the Tax Agency in the informal economy**.

¹ This study adopts a basic breakdown of the formal and informal economies, based on economic activity and the number of persons working at the local unit of activity:

- Informal economy: local unit working 10 workers or less or economic activity (NACE Rev. 1) = (1, 5, 14, 18, 19, 22, 29, 31, 36, 37, 45, 50, 51, 55, 63, 67, 70, 72, 74, 91, 93, 95)

- Formal economy: Others

4. Impact of the use of administrative records on indicators

We will now study the **potential impact of using administrative files on the basic indicators** produced from the Living Conditions Survey. Where possible, this simulation will attempt to replace the survey data with the data from the administrative file. If this substitution cannot be made, the original survey value will be left. No records are eliminated.

The basic indicators of the LCS based on household income are of two types: firstly, indicators measuring the **distribution of income** (relative poverty rate, Gini coefficient, etc.) and, secondly, indicators based on **level of income** (average income, poverty threshold, etc.). This report will analyse the impact of using administrative records on the relative poverty rate (broken down by age brackets) and on the average income per consumption unit.

Table 3 contains the indicators, replaced by the different sources of income. The first column contains the original survey results, together with the 95% confidence intervals. We then take the value of social benefits obtained from the Social Security system and recalculate the indicators.

The last two columns incorporate information from the Tax Agency, taking investment income and earnings from salaried employment and self-employment (in the case of self-employment, we take the maximum of the amount recorded in the survey and the amount indicated by the Tax Agency). The last column calculates the indicators using the methodology of the maximum amount for earnings from salaried employment in the informal economy.

Table 3: Impact of the use of administrative records on indicators (poverty rate and average income per consumption unit)

				With soc. benefits (Soc. Sec.)	With soc. benefits (Soc. Sec.) and investment income, self-employment (maximum) and salaries (Tax Agency)	With soc. benefits (Soc. Sec.) and investment income, self-employment (maximum) and salaries (maximum in informal economy) (Tax Agency)
		Confidence interval (95%)				
	Survey value	Lower end	Upper end			
Poverty rate						
Total	19.7	18.3	21.1	19.7	19.6	19.9
Under 16	23.4	19.9	26.9	23.8	24.6	24.4
16 to 64 years	16.8	15.5	18.1	17.0	16.9	16.8
65 years and over	28.5	25.4	31.6	27.0	26.1	28.3
Average income per c.u.	13,613	13,293	13,933	13,674	14,202	14,539

The table above shows that:

- If social benefits from the Social Security system are included, the relative poverty rate of older people is reduced, since the amounts in the administrative file were higher on average. The reduction is not significant and remains within the confidence interval.

- If we also take earnings from salaried employment and, in the case of earnings from self-employment, we take the maximum of the amount recorded in the survey and the profit declared to the Tax Agency, the situation is close to the original one.

- Lastly, a more refined methodology is used for earnings from salaried employment, making a distinction between the formal and informal economies. For the formal economy, the data is taken from the administrative file and, for the informal economy, the maximum is taken from the administrative file and the survey data. The indicators are closer to the original ones.

- The exception is average income per consumption unit, which, given that the recording of earnings progressively improves, increases with the change in methodology before finally obtaining a significantly higher value than the original one.

5. Conclusions

In this paper, we present the preliminary studies on the analysis of the linking of information on household income from the Living Conditions Survey and from data contained in administrative records.

For each component of income, we observe a specific case mix both in the comparison of the income amounts and in the classification of the income recipient.

In the calculation of the basic indicators using administrative sources, we see that **the use of administrative records does not appear to have a significant impact on indicators based on distribution of income**. However, it does have an impact on indicators based on income level as it significantly increases their value.

6. References

- [1] Regulation (EC) No. 1177/2003 of the European Parliament and of the Council of 16 June 2003 concerning Community statistics on Income and Living Conditions (EU-SILC). Official Journal of the European Union (Law), Vol. 46, No. 165 (3 July 2003).
- [2] INE. Encuesta de Condiciones de Vida. Metodología. www.ine.es
- [3] INE. Encuesta de Condiciones de Vida. La pobreza y su medición. Análisis de la renta y el gasto de los hogares. www.ine.es