

OFFICIAL STATISTICS

The forthcoming reform of the Spanish Living Conditions Survey: Extension proposals from an applied perspective

Jorge Onrubia

Universidad Complutense de Madrid (ICEI - UCM) and FEDEA

Abstract: The Spanish Living Conditions Survey (ECV) is the statistical production for Spain of the European Statistics on Income and Living Conditions (EU-SILC), encouraged and coordinated by Eurostat. This article presents some suggestions of improving the Living Conditions Survey elaborated by INE, in view of its upcoming reform, foreseen within the updating process of the EU-SILC project promoted by Eurostat. In particular, this paper reviews the incorporation of information from administrative registers, especially data from tax sources, as well as others from Social Security records. This work includes a series of proposals to improve the quality of currently gathered information in the ECV for Spanish personal income tax (IRPF), social contributions and social benefits. In addition, some extensions are proposed in relation to information on access to the main dwelling and its financing.

Keywords: ECV, EU-SILC, Household Surveys, Microdata, Tax Register Data

MSC: 62P20, 62P25, 68P01

1 Introduction

The European Union Statistics on Income and Living Conditions (EU-SILC) is a survey aiming to collect timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions, as well as other socioeconomic aspects such as housing, health, education, and benefits Eurostat (2020a). This instrument is anchored in the European Statistical System (ESS).

The EU-SILC project was launched in 2003 on the basis of a "gentlemen's agreement" in six Member States (Belgium, Denmark, Greece, Ireland, Luxembourg and Austria) and Norway. The EU-SILC database was released for the first time in 2004 for the EU-15 (except Germany, the Netherlands, the United Kingdom) and Estonia, Norway and Iceland. For a full description of the EU-SILC country-coverage please consult the implementation graph in Eurostat EU-SILC website.¹

¹<https://ec.europa.eu/eurostat/documents/203647/203704/EU-SILC+implementation+by+country.pdf>

The Spanish Living Conditions Survey (Encuesta de Condiciones de Vida, ECV) is the statistical production for Spain of the European Statistics on Income and Living Conditions (EU-SILC), encouraged and coordinated by Eurostat, the European Union Statistics Office. Therefore, the ECV belongs to the set of harmonised statistical operations for European Union countries. The fundamental objective pursued by EU-SILC national sections is to have a comparable source of reference on statistics on income distribution and social exclusion in the scope of the European Union (extended to other countries of the continent, such as Norway, Iceland or Switzerland).

The aim of this article is to present some suggestions to improve the Living Conditions Survey prepared by INE, in view of its upcoming reform, foreseen within the updating process of the EU-SILC project promoted by Eurostat. The content of this article was presented on 6 September 2019, in the round table organised by the Spanish Statistical Office (Instituto Nacional de Estadística, hereinafter, INE), within the XII Public Statistics Conference (Jornadas de Estadística Pública) held in Alcoy.

Essentially, this article reviews the potential that, in my opinion, the incorporation of information from administrative registers may have, particularly data from tax registers. The quality of these registers and the extensive experience in their maintenance provide a great opportunity to enhance a great statistical product such as the ECV produced by the INE since 2004.

2 The possibilities of using administrative records in ECV

A comprehensive guide to the possibilities that can be offered by incorporating individualized statistical information from administrative registers (corresponding to population, taxes, social security data, public social benefits, and health and education activities, as main areas) can be consulted in J'antti et al. (2013). As highlighted in this report, "In the specific context of social statistics, the re-use of existing data and in particular administrative data has been identified by the European Statistical System as a key developing area in the process of modernising and streamlining social surveys". However, we have to take into account the different degree of use of administrative data by countries. In some cases it depends on the quality and development of administrative records. In other cases, it depends on the legal limitations related to the dissemination of personal information. The importance of the availability and usability of such data by national statistical offices, including timeliness of receipt and comparability, should also not be overlooked. Finally, an issue that is sometimes more difficult to achieve is the cultural change both in public administrations that provide the data and in the national statistical offices.

However, between countries, there are still notable differences on how the information is obtained to build the national databases to be integrated into EU SILC, as well as in their degree of intensity when combined. Even in direct surveys there are differences. Most countries carry out face-to-face interviews by Paper And Pencil Interviewing (PAPI), while other use Computer Assisted Personal Interviewing (CAPI). In Germany, the interviews are usually self-administered by respondents. The use of administrative records has been a widely used source of information in the Nordic countries for decades, as well in the Netherlands and Slovenia, and to a limited extend also in France, Ireland and Latvia. Conducting telephone interviews to supplement information obtained through the use of administrative sources is also a method used in some countries. In Spain, as explained José María Méndez in his article in this same volume of the Spanish Journal of Statistics (Méndez, 2019), in the Spanish ECV, since 2005, information has been collected through CAPI, with a pilot test having been carried out in 2017 (which is expected to continue in 2019), in which the so-called "multi-channel collection" CAWI-CATI-CAPI is being trialed. With the modernisation of data collection, greater

efficiency in ECV production is pursued, although it is recognised that this requires major challenges to be faced, given the complexity of ECV itself, and of the information it incorporates.

On consistency over time and on comparisons between countries in EU SILC when combining data provided directly from the respondents' questionnaires with data obtained from administrative records, you can see Krell et al. (2017). This article shows that achieving successful administrative data usage is not an easy task, and that methodological transparency is essential for reaching such an objective.

In the case of Spain, in the Living Conditions Survey 2013, a new methodology was adopted in the production of data regarding household income, consisting of the use of administrative files, although the information available in the questionnaires continues to be used as well. In my opinion, the incorporation of information from administrative registers implies an improvement in the quality of the data. The use of these administrative sources, given the high quality of the statistical records of the Spanish public administrations, represents a significant efficiency gain in the data collection method.

Although the first wave of the ECV that incorporated administrative-source income was that of 2013 (annual income corresponding to 2012), since the Tax Identification Number (NIF) had been available since 2009 (with 98% coverage), retrospectively data have been provided with this source for the waves from 2009 to 2012 (with annual income, respectively, for 2008 to 2011). This duplication of information in those waves has been useful in order to be able to comparatively analyse the differences between both sources and their impact on the measurement of inequality and poverty, among other issues.

Initially, the objective pursued was to enhance the production of household income variables. In this way, household income related data are prepared using a mixed methodology, which combines the information provided by the respondent with the administrative records of the State Tax Administration Agency, Social Security, the Tax Office of Navarra and the Diputaciones Forales of Bizkaia and Gipuzkoa. It is worth noting that the Diputación Foral of Alava did not participate in this contribution of information. In this case, it is simply a question of lack of willingness, under a peculiar interpretation of tax autonomy.

3 Proposals for the extension of the ECV

This section of the article incorporates some proposals with possible extensions of the information currently contained in the ECV, which could be developed mostly from the administrative sources (from tax and social security administrations) already used since the methodological change of the 2013 wave.

3.1 Annual ECV extension modules

The ECV incorporates an annual module pre-defined by Eurostat. These modules complement the basic information coordinated according to Eurostat's EU-SILC project. These ad-hoc modules are developed each year and complement the information provided by the variables permanently collected in EU-SILC, with supplementary variables about unexplored aspects of social inclusion Eurostat (2020b). Since 2005, which included a module dedicated to the Intergenerational transmission of poverty, the ECV has incorporated the following modules: (2006) Social participation, (2007) Housing conditions, (2008) Over-indebtedness and financial exclusion, (2009) Material deprivation, (2010) Intra-household sharing of resources, (2011) Intergenerational transmission of disadvantages, (2012)

Housing conditions, (2013) Wellbeing, (2014) Material deprivation, (2015) Social/cultural participation and material deprivation, (2016) Access to services, (2017) Health and children's health, and (2018) Material deprivation, well-being and housing difficulties. For the ECV 2019, the module established is Intergenerational transmission of disadvantages, household composition and evolution of income. As it can be seen, chosen themes are often recurrent, although variations are introduced.

One possibility, in my opinion, of great value, would be to offer a three-year module linked to the subject matter of the three-yearly Encuesta Financiera de las Familias (EFF) (Bank of Spain, 2020), drawn up by the Bank of Spain and included in the Spanish National Statistical Plan.² The main objective of this module would be to enable consistency analysis between the ECV and the EFF, as well as to carry out complementary studies on financial and savings decisions of Spanish households. The design of the module should take into account that the reference year of the annual income collected in this wave of the ECV (that of the previous year), from the personal income tax (IRPF) records, would coincide with the reference year of the EFF.³

With regard to the information to be incorporated, the questions asked of the interviewed households should be related to savings (proportion of income allocated to savings), indebtedness (level and destination) and the composition of the portfolio of assets that make up their wealth. As background, we can mention the ECV 2008 module aimed at "Over-indebtedness and financial exclusion", and the one planned for ECV 2020, on "Over-indebtedness, consumption and wealth".

An experience on ad-hoc modules of the EU-SILC with the incorporation of information on consumption and wealth of households, with an orientation to the testing of the HBS and the HFCS, is the one developed by the Statistical Institute of Belgium for the SILC of Belgium (Statistics Belgium, 2018). The experience, as recognized in the assessment study is the high non-response in the case of consumption questions, the underestimation in the case of wealth questions, largely attributed to the breadth and little specificity of the questions. It may also be justified by disincentives for households to transmit reliable information (which we also know it takes place in the HFCS).

3.2 Broadening the information contained in the ECV from tax records

As mentioned above, there is now a broad consensus about the enormous potential role of administrative records in enriching the information provided by household surveys, especially tax records. According to the coordination guidelines established by Eurostat, from 2013 onwards, the information on household income provided in the national surveys on income and living conditions integrated into EU-SILC will come from tax records contained in the files of tax administrations for the management of personal income taxes.

The use of these tax sources has also made it possible to incorporate administrative information on personal income tax (IRPF) payments into these surveys. Likewise, the administrative records of the Social Security allow the incorporation into the ECV of the amounts corresponding to the social contributions paid by employees, and self-employed. Specifically, the Spanish ECV incorporates, in each wave and for each member of the household, the total amount of withholdings charged and payments on account made for IRPF and Social Contributions withheld or paid during year $t - 1$ (the year for which the annual income is offered in the ECV). In addition, the amount, positive or negative

²The Spanish Survey of Household Finances (EFF) is part of the Eurosystem's Household Finance and Consumption Survey (HFCS), coordinated by European Central Bank, which collects household-level data on households' finances and consumption (European Central Bank, 2020). The HFCS datasets from the first and second wave were released respectively in April 2013 and December 2016.

³The last wave released by the EFF has 2017 as reference year (Bank of Spain, 2019).

(depending on whether it is an amount paid or a refund received), of the differential tax liability for year $t - 2$ is also incorporated as variable.

However, with the complete information from the IRPF and Social Contributions records, we think it would be useful:

1. To include a variable for the “differential tax liability” of IRPF corresponding to the fiscal year $t - 1$ (the reference year of the annual income, and the withholdings and payments on account of the IRPF). This would make it possible to have the “total tax liability” of the IRPF, corresponding to that reference year. Nowadays, users have a “total tax liability” of the IRPF, approximated by a cash criterion: withholdings charged ($t - 1$) + payments on account ($t - 1$) \pm differential tax liability ($t - 2$).
2. To offer separate information on the amounts corresponding to the total amount of employee/self-employed made for IRPF during year $t - 1$. This information is provided by the State Tax Administration Agency (AEAT), and by Basque Country Provincial Tax administrations, and Navarre Foral Tax Administration). And, on the other hand, the amount of employee/self-employed Social Contributions provided by the Social Security. Currently, variable HY140G includes the sum of IRPF’s withholdings charged and payments on account for $t - 1$, and Social Contributions paid in $t - 1$, while the differential tax liability for $t - 2$ is provided in variable HY145N.

Likewise, it would be advisable to improve the information referring to income from personal entrepreneurial economic activities, identifying the estimation regime applied in the IRPF (direct estimation of incomes and expenditures, objective estimation by means of modules, and objective estimation by coefficients for farmers and stockbreeders). Also, if it is the case, it would be interesting to reflect whether the data included in the survey comes from the tax return or from the response to the interview form. In relation to this question, it would be very useful for studies on compliance and tax evasion, to ask in the interview questionnaire for the amount obtained from this type of income, in order to compare it with the income reported in the IRPF return.

3.3 Detailed information on Spanish personal income tax (IRPF) in the ECV

In the same line, it would be desirable to incorporate, in the ECV’s file of members, more detailed information on IRPF variables, available in the annual IRPF’s withholding statement of household members. In this case, from the perspective of the ECV user, the objectives would be:

1. To allow greater precision in the analysis of the effects of taxation on personal income;
2. The possibility of using ECV, exploiting the richness of its socio-economic variables, to carry out micro-simulation exercises for alternative IRPF designs. Within this field, this extension of IRPF variables would be very useful to provide greater precision and reliability to the EUROMOD tax “modules”.⁴

A proposal of IRPF’s variables to be incorporated could be the following:

- Variables of primary interest: type of tax filing (Individual, Joint taxation, only withholdings); general taxable base and savings taxable base; amount of minimum personal and family allowances; State/Autonomous Communities gross tax liabilities; State/Autonomous Communities tax credits; total tax liability; differential tax liability; total refundable tax credits.

⁴EUROMOD is the open-access Tax-Benefit Microsimulation Model for the European Union, for the European Union. EUROMOD allows researchers and policy analysts to calculate, in a comparable manner, the effects of taxes and benefits on household incomes and work incentives for the population of each country and for the EU as a whole. For more information, see EUROMOD website: <https://www.euromod.ac.uk>

- Other variables with detailed information: amounts of income components by source (labour, financial and non-financial capital incomes, rents, business and professional incomes, capital gains), contributions to pension plans and other social protection systems, details of personal and family allowances, etc.

3.4 Social benefits incorporated in ECV

Another interesting possibility for extending the information offered by the ECV concerns the exploitation of administrative records containing data on social benefits. Firstly, for all benefits, it would be very useful to provide differentiated information on the public or private nature of the entity paying them. At present, this is not the case for benefits of an educational kind.

Another change that would be quite useful, from the analysts' point of view, would be the differentiation of severance payment and unemployment benefits. Both are now included in a single variable, and estimates must be made according to the legal limitation on the amounts of the latter.

Other possibilities to enrich currently provided information in the ECV could be:

1. Incorporation of information on the type of unemployment benefit, depending on whether it is a contributory benefit or family aid for the long-term unemployed. It would also be interesting for the analysis of this benefit to be able to identify the receipt of the subsidy in a single payment.
2. In the case of educational transfers, it would be convenient to incorporate some information on the educational level at which the aid is aimed (childhood education, primary, secondary, high school, and university education).

Likewise, it would be very useful for the territorial analysis of these benefits to be able to differentiate in one variable the Public Administration that makes the payment (State, Social Security, Autonomous Communities, Local Corporations).

3.5 Information on housing acquisition, finance, and rental

The dwelling in which households reside is a field of considerable interest for the expansion of information in a forthcoming ECV reform. It should be borne in mind that, on the expenditure side, the Spanish Household Budget Survey (Encuesta de Presupuestos Familiares, EPP) does not provide information on the house purchase, on considering it to be an investment, nor on other secondary dwellings owned, either for the use of by household members, or for rental.

The incorporation of information related to the dwelling should come from different tax registers, such as those of the Real Estate Tax (IBI), IRPF, the Property Transfer Tax (Impuesto sobre Transmisiones Patrimoniales Onerosas, ITPO) or the Inheritance and Gift Tax. In addition, the information available in the public Real Estate Registries could be used.

Given the volume of information available in the field of housing, the objectives to be achieved with the extension of the ECV should be specified. Thus, in the first place, the inclusion of new variables should be aimed at improving the analysis related to the acquisition, tenure and financing of real estate by households. This is an improvement already planned for 2021 in the action plan of the Spanish National Statistics Institute (INE). Secondly, the information to be included must make it possible to identify whether the main dwelling under the ownership regime was a new construction (first transmission) or was a used dwelling (second transmission), as well as to record the year of construction and the year of acquisition by the household.

It would also be desirable to incorporate the purchase price of the dwelling, the mode of acquisition (purchase, inheritance, donation), as well as the Cadastral Value (gathered in the administrative registers of the AEAT).

As regards house finance, it would be very useful to identify whether there are any outstanding loans, and the amount pending repayment in the reference year. It would also be interesting to be able to differentiate the type of loan involved (mortgage, other loans with financial institutions, loans from relatives), as well as the term, and the interest rate applied.

When it comes to second housing, it would be interesting to incorporate information on other (non-main) dwellings available to the user or rented (the ECV now only has data on the main dwelling of the household).

For main dwellings that are rented, the ECV should incorporate detailed information on the amount paid for rent, differentiated from the rest of the residence expenses that are paid by the tenant. In the case of secondary dwellings, it would be interesting to include information, as well as the main reason and duration of the rental.

4 Concluding remarks

The purpose of this article, as stated in the introduction, was to set out some possibilities for improving the information offered by the Spanish ECV, essentially by incorporating data from administrative registers. The forthcoming reform of the EU-SILC project, promoted by Eurostat, may be a good opportunity to enhance the use of administrative records as a source of rich and reliable statistical information.

In Spain, the quality of these administrative registers and the long experience in their maintenance and continuous improvement can provide an important added value to a valuable and contrasting statistical product such as the ECV carried out by INE. However, I am aware that progress in this area requires sufficient human and material resources to be provided to Spanish Statistical Office (INE). In my opinion, undoubtedly, the widespread use of the ECV over the past 15 years, in practically all areas of applied social sciences, justifies the necessary investment.

Acknowledgments

An earlier version of this paper was presented at the XII Public Statistics Conference (Jornadas de Estadística Pública) (Alcoy, Spain, September 9, 2019). I gratefully acknowledge the useful comments and suggestions received from the participants in the abovementioned meeting. We would also like to thank Andoni Montes for his valuable comments and suggestions after a careful reading of the manuscript. All the remaining errors are entirely our responsibility. I acknowledge the financial support of Spanish Ministry of Science, Innovation and Universities (before Ministry of Economy and Competitiveness), Project ECO2016-76506-C4-3-R.

References

Bank of Spain (2019). Encuesta Financiera de las Familias (EFF) 2017: métodos, resultados y cambios desde 2014. Technical Report 4/2019, Boletín económico/Banco de España.

- Bank of Spain (2020). Survey of Household Finances. Technical report, Bank of Spain. https://www.bde.es/bde/en/areas/estadis/estadisticas-por/encuestas-hogar/relacionados/Encuesta_Financi/.
- European Central Bank (2020). Household Finance and Consumption Survey (HFCS). Technical report, European Central Bank. https://www.ecb.europa.eu/stats/ecb_surveys/hfcs/html/index.en.htm.
- Eurostat (2020a). European Union Statistics on Income and Living Conditions: Ad-hoc modules. Technical report, Eurostat. <https://ec.europa.eu/eurostat/web/income-and-living-conditions/data/ad-hoc-modules>.
- Eurostat (2020b). European Union Statistics on Income and Living Conditions (EU-SILC). Technical report, Eurostat. <https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>.
- Jäntti, Markus, Veli-Matti Törmälehto, and Eric Marlier (2013). *The use of registers in the context of EU-SILC: challenges and opportunities: 2013 Edition*. Publications Office.
- Krell, Kristina, Joachim R Frick, and Markus M Grabka (2017). Measuring the consistency of cross-sectional and longitudinal income information in eu-silc. *Review of Income and Wealth* 63(1), 30–52.
- Méndez, José María (2019). The Spanish Survey of Living Conditions (ES-SILC). *Spanish Journal of Statistics* 1(1).
- Statistics Belgium (2018). Preparation for the revision of EU-SILC : Testing of rolling modules in EU-SILC 2017 - Final Report. Technical report, Statistics Belgium.