

Using Tax Sources in Household Survey Sampling

Sébastien Faivre

At INSEE, household survey frames have typically been based on the population census (RP). However, INSEE has sought to develop an alternative survey frame in response to the following three factors:

- the complexity of the new continuous survey which, while providing fresher data, also adds new constraints that are not easily overcome in the case of random survey sampling, for example;
- the emergence of new demands (e.g. online surveys, “individual” surveys);
- the availability and quality of a new source: tax datasets.

INSEE therefore decided to explore the possibilities offered by the tax datasets provided by the Public Finance Directorate-General (DGFIP).

1. Tax Records Provided to INSEE by DGFIP [1]

All French citizens are taxed:

- on their income;
- on the residence¹ that they occupy on 1 January of each year. This tax is known as the residence tax (“taxe d’habitation” - TH);
- on the residence(s) that they own. This tax is known as the property tax (“taxe foncière”).

To do this, information regarding income, wealth and place of residence must be reported to the DGFIP. **A single declaration can be made for multiple persons** as declarations are made **by tax household**.

The term “tax household” refers to all persons listed on the same income declaration. **There may be multiple tax households in one home**: for example, an unmarried couple in which each member declares his/her own income counts as two tax households. Equally, **an individual can belong to multiple tax households**, which are “departmental”. Where a person lives and pays his/her income tax in department D1, and possesses a secondary residence, for which he/she pays the residence tax in department D2, he/she will belong to two different tax residences. This is the main difficulty that must be overcome in order to be able to construct a survey frame of “individuals” without double-counting based on tax databases.

The tax household is therefore a concept close to that of the home (in sociology), family home or economic household, apart from a few specific features. As such:

- each person in a de facto union is counted as single (except for the wealth tax, which is taxed collectively), with, where applicable, the inclusion of unmarried children aged under 18 in the custody of one or other individual;
- the income of individuals in a civil partnership (PACS) has been taxed together since the introduction of the Finance Act of January 2005, unless the PACS is ended during the year after its signature;
- a monastery or convent is considered a single tax household.

On the other hand, **the same individual may make multiple declarations of the same type, or no declarations, as the case may be**.

The DGFIP is responsible for collecting and monitoring these declarations, and subsequently calculating the amount and collecting tax for each household.

For each year N, the DGFIP has computer databases containing information on the following declarations:

- income for year N-1 for all tax households in France;
- the individual composition of tax households and their address on 1 January of year N;

¹ and is levied on owners, renters or occupants free of charge.

- the description of completed habitable dwellings (including the address) in year N.

INSEE receives these databases at the beginning of year N+1.

It should be noted, however, that a small number of individuals are still not included in the tax records. These are mainly individuals who have recently reached adult age (mostly young people aged 18, but this factor can persist until age 22). In some cases, because these young people no longer entitle declarants to tax reductions, they may not be reported as being in the tax household of their parents. This will have no impact on the amount of tax payable by the parents. Furthermore, as young people of adult age are not compelled to make tax returns if they do not have a income or their own place of residence with the associated residence tax liability, they do not appear in the aforementioned files.

The information contained in these files for children aged under 18 is minimal, because only the date of birth is available for these individuals. Children in a joint custody arrangement also appear on two separate tax returns. While it is possible to identify them independently in both households, it is impossible to identify the two households that relate to the same child.

2. The Strengths of the Tax Source as a Survey Frame

Every year, INSEE seeks to use the tax source to construct two exhaustive frames: one for dwellings and one for individuals.

A large amount of methodological work has been undertaken to identify the primary residences as defined by INSEE from among the dwellings listed by the tax authorities, and to identify instances of double-counting of individuals known to the tax authorities and grouped together on the personal tax database.

These frames may be used naturally as survey frames for sampling in household, housing or individual surveys, where they offer the correct properties required from a survey frame, after statistical processing:

- completeness
- freshness
- no duplicates
- presence of auxiliary information (e.g. household composition, age of occupants, income variables, etc.).

3. The Use of the Tax Source as a Survey Frame for Household Surveys

3.1. Employment Sampling in the 2006 Residence Tax [2]

The employment survey uses random sampling: a certain number of “small zones” are selected, known as areas, or clusters, and all dwellings in the selected areas are surveyed; dwellings are by design spaced as close together as possible. Specifically, continuous employment survey sampling is based on samples of 3,200 sectors and 120 dwellings: each sector is divided into six clusters of 20 dwellings surveyed consecutively, with each cluster surveyed for six consecutive quarters. The current sample was taken from the tax source (residence tax records for 2006) and was gradually rolled out from 2009 onwards.

3.2. Some Households are Already Drawn from the Tax Source

The population census cannot be used to meet all sampling requirements, insofar as it only enables sampling of dwellings (information on the identity of the dwelling occupants is not available) and, in the case of “paper” respondents to the population census, the address can only be obtained by scanning the address box of the household questionnaire. Because of this, the tax source must be used in the following cases:

- one-on-one household surveys sampling individuals (as is the case for PIAAC² 2012 and Training and Professional Qualifications 2015);
- one-on-one household surveys sampling dwellings by using tax data in the survey frame to over-represent certain sub-populations (Wealth survey) or certain land uses (oversampling for New Priority Districts in the 2016 Quality of Life and Security survey);
- multi-modal surveys requiring mass-mailing (Daily Life and Health survey, pilot online surveys) or a telephone number search via AnnuCapt (Information and Communication Technologies survey and Camme monthly business survey).

3.3. Planned Redesign in Favour of Single Tax Source Sampling from 2020

In late 2015, INSEE's Management Committee launched the Nautile project with the aim of introducing single sampling for all INSEE household surveys using the tax source by 2020.

The expected gains of the project relate in particular to:

- improved accuracy for a given sample size by returning to an exhaustive survey frame and discontinuing rotation of the population census survey frame
- a single application used to meet all sampling requirements
- more straightforward fieldwork planning.

References

[1] Olivier HAAG, « La source fiscale : une base de sondage qui s'impose pour les enquêtes auprès des logements et des individus », Symposium de Statistique Canada 2013.

[2] Vincent LOONIS, « La construction du nouvel échantillon de l'Enquête Emploi en Continu à partir des fichiers de la Taxe d'Habitation », *actes des journées de méthodologie statistique 23, 24 et 25 mars 2009*.



Department of statistical methods
Version 1, published 10 October 2017

² Program for the International Assessment of Adult Competencies, survey on adult literacy and numeracy skills carried out within the OECD framework.