

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

STATISTICAL OFFICE OF THE EUROPEAN UNION

**Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses
(The Hague, The Netherlands, 10-11 May 2010)**

Working paper 8
6 May 2010

Topic (iii) of the provisional agenda

FIT FOR THE CENSUS

The french use of housing tax in the new census method

Note by the National Institute of Statistics (INSEE), France*

SUMMARY

1. France has implemented, since 2004, a rolling census: a survey is conducted annually on a portion of the territory. At the beginning of each year, about 14% of the population completes a census form. At the end of each year, an official population (used in many financial and regulatory areas) is calculated, for each *commune*, based on the last 5 surveys. This “legal population” is calculated for each *commune* (of which there are 36,682 in France) regardless of the survey’s date in its territory.
2. The method is mainly based on population surveys. But an administrative source, the housing tax, is used to put back to the same reference date data collected within different years. This paper aims to present the details of this "statistical" use of administrative data.
3. When the survey in a *commune* under 10 000 inhabitants has taken place one or two years before the reference date of the census, the number of homes subject to the tax allows to extrapolate the population. The administrative data is used as a trend factor -not a level factor-, to refresh the data from the last survey. (Part 1)
4. The first step is collecting data from the tax administration. The data is constructed according to rules and practices that meet the fiscal targets. We had to understand their way of elaborating and schedule update. A second phase of expertise then led to the conclusion that the separation between occupied dwellings and other dwellings (seasonal or secondary used, vacant),

* Prepared by François Clanché, Head of the Demography Department.

was not sufficiently reliable in the tax databases. So, the final method takes in account the evolution of the total number of dwellings from one year to another. A correction coefficient, from the figures collected during the previous census, takes into account the trend in the number of persons per dwelling in the town.

5. Finally it was necessary to establish "correction" rules in the very few situations where the data changes from year to year is clearly erroneous, with significant decreases one years followed by increases of similar magnitude following year. (Part 2)

6. Finally, this process has been operated for 2 years (legal populations have been computed and published in December 2008 and December 2009), and has raised no objections, especially from the 21 000 municipalities whose population has been calculated by this method. It however requires significant permanent human resources. (Part 3).

I. WHY USE THE HOUSING TAX ?

A. The necessity of additional data to complete surveys

7. With the French method of census, the 36 000 *communes* with fewer than 10,000 inhabitants are surveyed exhaustively, rotating once every five years. Their populations are however calculated every year, like all *communes*. For reason of statistical consistency, but also regulatory fairness, it is necessary that the "population legal figures" of all the *communes* are calculated for the same year, called "base year".

8. For reason of statistical robustness, it was decided that this base year is the middle year of the five-year cycle of collection. For example, for populations published at the end of the year 2009, based on surveys conducted between 2005 and 2009, the base year is 2007. For the fifth of small communes surveyed in 2007, the result of the investigation is exactly taken into account. For the municipalities surveyed in 2008 or 2009, we assume a steady growth between the last published figure (which had 2006 as reference) and the most recent survey (2008 or 2009) to project the year 2007 by linear interpolation.

9. However, for the communes recorded in 2005 or 2006, we must "refresh" the data to determine a figure in 2007. A simple extrapolation of data on one or two years, using the trends from the last two surveys, was possible. But it involved significant risks. Indeed, these municipalities are small or very small (half of them have less than 400 inhabitants). It is therefore not certain that each year falls within the "trend" of previous years: the "large numbers law" does not work on such low frequencies. In a town of 1,000 residents (about 500 dwellings), there can be 5 new dwellings built (subdivisions, small building) each year for two years, then no other construction during the following 2 or 3 years. A simple extrapolation would lead to misdiagnosis, with an over or under estimation of the population of several points.

10. Before the implementation of this new method, « additional censuses » were organized, between two general censuses, to take in account important population expansion in small communes. As increasing *communes* are more numerous than decreasing ones, it was necessary, in the new device, to keep a way of recording evolutions between 2 surveys.

B. Advantages of the housing tax

11. The 2002 law that establishes the new census method explicitly provides the opportunity to complete the survey data with information from administrative sources.
12. Therefore we've searched for source that can provide guidance on annual population growth, reliable at the level of each *commune*. There is currently no administrative database including all residents with their addresses in France. However, there is a good quality source on the number of dwellings: the file of "housing tax" (HT).
13. Any occupant of a dwelling has to pay this tax each year. Even if he holds several homes during the year, he pays the tax only once. Although collected by the state, the proceeds of the tax are paid entirely to local authorities, and primarily the *communes*, where it represents a major source of funding. The *communes* are financially interested in the quality and completeness of the database. In addition, they are regularly consulted by the Tax administration to check if all the premises are included. This is an advantage for the statistician, because it's a pretty good guarantee of completeness, and this allows municipalities to set their liability in case of dispute. The source also allows to distinguish -at least in principle- permanently occupied dwellings (principal residence) from second homes or casual vacancies.
14. Moreover, as people rarely want to be unduly taxed twice, it is unlikely that include double counting.

C. The operating principle originally envisaged

15. The principles originally planned to use this source were as follows:
Through observation of changes in the number of primary residences subject to the tax you calculate the rate of growth in the number of dwellings between the years of the last census survey and the reference year. As example 1000 homes in HT 2006, 1010 in 2007, would be a $10/1000 = 1\%$ growth per year.
16. As the average number of persons per primary residence is not stable over time, we calculate the trend in the number of persons per occupied dwelling between the last two census surveys conducted in the municipality. For example, - 0.5% per year.
17. It is concluded that overall, between 2006 and 2007, the population has increased in all likelihood ($1 - 0.5 = 0.5\%$) per annum. This coefficient is applied to the 2006 figure to arrive at the 2007 figure.
18. The housing tax is used in this schema in evolution, in addition to the census, but not level: the number of housing of TH is not replacing the RP. These data will then be "confirmed" (or partially undermined) by the census survey involved one or two years after the calculation: their impact will be limited in time.

II. THE ACTUAL ESTABLISHMENT OF THE DEVICE

19. Once such principles, we must implement the "real" data from the tax. An agreement was signed between INSEE and tax administration, made possible after consultation with the CNIL (national commission data processing and freedoms). Each year, the tax administration provides INSEE a file of all dwellings covered by the HT.

A. What premises should be taken into account?

20. The file received by INSEE is built by the tax authorities. It contains 42 millions records. It is itself the result of the merger of two bases:

- A "land file, built primarily from the cadastre, which essentially gives a description of "physical" space (exact address, type of room, size).
- A "taxpayer file", including data from income tax, which provides information on taxable persons (name, number of persons, marital status, address of the declarer).

21. These two databases are merged by the tax authorities. Some premises are found only in one of two bases, which is itself often an indicator of their nature or quality.

In the database, all local units are not "housing units" as defined in the census. It contains also:

- Dependencies (garages, room service, swimming pools, shops...) located in close proximity to housing. In most cases, we can identify those premises by the combination of descriptions "physical" (from the land base) and their information "taxpayer" (they have the same registrant).
- Furnished premises for use as businesses or associations. In most cases, information on the nature of the declarer leads to isolate this situations, because they are not ultimately subject to the housing tax.
- Buildings used by rural farms (barns, stables...). Here, information on the nature of the space allows them isolated.

22. There is no variable in the file that exactly matches the statistical concept of "housing units". INSEE built a "filter", using several variables, sometimes from "land" origins, sometimes from "taxpayer", to reduce the file only to likely local residential locals. This filter leads to remove 25% of premises originally present in the tax file.

23. INSEE constructed therefore, for its use, a file called the "filtered HT" to be used in several operations related to the census: updated inventory of buildings located (RIL) in large towns, quality control of survey gathering in small *communes*, calculating the population by extrapolation in some small *communes*. The following text will only cover the third use.

B. The concept of primary residence

24. Among the identified units, only the "primary residences", occupied the major part of the year, should in principle be taken into account to participate in the "Update" census data. Despite the existence of such a category in the HT source, it is not easy to spot in this file, and this for several reasons.

25. Is considered "primary" for the housing tax a residence where a person has declared his income (under the Income Tax Act) at the beginning of the previous year. But people do not necessarily declare their incomes where they live most of the year: it is not required. The housing tax rates vary by commune. They are also different depending on whether a dwelling is declared or not "primary residence". Thus a person may have an interest to declare his income in a home where she lives only a small part of the year. It may, therefore, in level, imply significant differences between the number of primary residences under the census and the same number as defined in the HT. Most of the time, "secondary residence" pay more tax than "primary residence": neither the *communes* nor the tax administration are interested in reducing their number.

26. In addition, the one-year lag between the declaration of income tax and taking into account the property tax makes analysis difficult and changing from one year to another, especially on small numbers. A change in status may also result from a "tax strategy" of a taxpayer who has an interest in changing residence within the meaning of the tax without actually moving.

27. Empirically, there is greater volatility of the series of "primary residences" as a serie of housing in general: strong evolutions year on year, significant reversals of trends are much more significant in series of primary residences than in the general series of dwellings.

28. It soon became apparent, when the first tests of applications of the method originally planned that these irregularities resulted a significant number of highly improbable developments of the populations concerned. For all the data to be credible, "case by case" managing was necessary in a large number of situations. It would have been expensive, would certainly have created distortions and inequities between *communes* and could have caused disputes and even litigation. If, in the coming years, the *communes* make a better work on their tax bases, the quality on this series will increase and we will perhaps return to the first projected solution.

C. The solution finally chosen

29. It was therefore decided to implement a more robust device: to set the trend extrapolation on one or two years, the total number of dwellings is taken into account. It is corrected by the trend of the average number of persons per housing (and not the average number per primary residences). This solution also has the advantage that the number of dwellings is better known by the *communes* than primary residences. It is less controversial in the *communes* because it is not sensitive to "tax strategies" of residents.

Example : Monchiffray, fictitious *commune* with census in 2005

	1999	...	2005	2006	2007
Census Data :					
Number of dwellings	858		924		
Number of inhabitants	1859		1950		
Mean number of inhabitants by dwelling	2,17		2,11		
Housing tax data :					
Number of dwellings			910	923	951

30.	Average annual evolutions according to censuses of 1999 and 2005 :		
	Of the number of dwellings :	1,24%	(1)
	Of the population :	0,80%	(2)
	Annual evolution difference ((1) - (2))	0,44%	(3)
	Average annual evolution of the number of dwellings according to the HT between 2005 and 2007 :		
		2,23%	(4)
	Estimated average annual growth of the population (4 - 3) :		1,78%
	Estimated population for 2007 :	2 020	inhabitants

D. Management of aberrant developments

31. For a given commune, the sequence of numbers of housing is in most cases a regular pattern: during a period of 3 years, 83% of *communes* have increased three years in a row, 9% are in continuous decline. For about 7% the profile is irregular but without significant reversals. Thus, the use of developments in one or two years does not pose a problem.

32. However, for 0.5% of the *commune*, patterns are surprising: a year of "peak" followed by a return to a regular pattern, or one year of "hollow" followed by a return to trend, or double inversion.

33. We had to set thresholds, depending on the size of the town, from which it was decided to "smooth" artificially the series, assuming that the year of peak (or trough) was due to errors. This threshold was set at 6% of homes in a commune of 100 inhabitants, 4% per 500 inhabitants, 0.6%

34. Each year, 300 municipalities (out of 14,000) are thus subject to special treatment. These corrections are systematic: the same treatment is applied to all municipalities with the same "profile". These apparent anomalies in the files are subject to an alert to the tax administration, but unfortunately we do not always know whether the corrections have been made thereafter.

35. Each year after the date where files are sent to Insee, the *communes* may report errors to the tax authorities. But the corrections are not taken into account in a centralized database. Thus, we have no knowledge of these corrections. This is unfortunate, because the correction of the year N-1 improves the analysis of trends.

III. AN ANNUAL OPERATION EFFECTIVE BUT EXPENSIVE

36. All these management and treatment rules have been constructed and developed over the years from 2005 to 2007, for a first implementation in full scale in 2008, with at the first calculation of the populations of all *communes* based on surveys from 2004 to 2008. The process is now repeated every year.

A. Centralized checking and corrections

37. Each year, the files relating to the housing tax, corresponding roughly to the situation at the

beginning of the year, are made available by INSEE in December. Specifically, 107 files are sent by central tax administration, corresponding to the 107 tax centers in France.

38. INSEE receives the file, checks their integrity and the likelihood of volumes, and then performs all the automated processes described above: filtering, systematic corrections outliers. These transactions are added to other treatments associated with other uses of TH, including addresses. This takes about two months, leading to provisional estimates.

B. Local expertise on particular issues

39. The calculations, performed centrally, are sent to the regional directorates of the INSEE for checking. The review, during the first campaign of calculation by *commune*, showed, despite the checks already mentioned, some strong changes from one year to another in *communes* yet seemingly "stable". The analysis of such situations has usually revealed the following phenomena: the appearance-or, disappearance - of an important address in the commune linked to the timing of administrative procedures, which is not consistent with the definitions or protocols of census.

40. Two examples:

A building of 20 homes, dilapidated, called to be destroyed, uninhabitable because walled, disappears from the housing tax file between 2005 and 2006 because it was effectively destroyed. But in January 2005, the enumerator did not take into account the building, following instructions, since it was completely uninhabitable. There is therefore a risk of reducing the population of the town by about 40 people in 2006 by subtracting unfairly from the population enumerated in 2005. In this case, the technical solution is to remove the building from the TH file 2005.

41. Conversely, 25 units appear in the HT file of a small town between 2005 and 2006. At the same address, a retirement home was surveyed in 2005 as a community, with 29 people. In the meantime, the status of the community has evolved and its inhabitants are now eligible to tax, which was not the case in 2005. It would be wrong to add these 25 units when calculating the evolution from 2005 to 2006.

42. These specific situations, very few in number, can only be detected and analyzed in a case-by-case basis, by the regional teams. The national team "prepares" however the work, by identifying cities whose changes in HT are quantitatively significant and where one address has a significant impact. This allows targeting expertise.

C. A technical and political success

43. In 2008 and in 2009, this calculation method was applied to two-fifths of "small towns", about 21,000 municipalities in total, which are not the same every year. Half of them are extrapolated over a year, another two years.

44. In 2008, first year of publication, when the municipalities received the letter informing them of their population and how they had been calculated, hundreds of them asked for further explanation. But none have disputed the result, and there is no official dispute in this area. This favorable response has been made possible by an important preliminary communication process: the

INSEE regional officials and the teams in charge of the census met the mayors during the years 2007 and 2008, and explained the principle of this calculation and its substantiation.

45. In 2009, a second set of calculations was performed, covering the base year 2007. Developments in relation to the previous point have been reported. Afterwards, for communes that were surveyed in 2009 (after having been in 2004), the quality of the "prediction" by the housing tax data for 2006 has been measured. Overall, this second year has not revealed any unanticipated problems, and has not led to question the method. Even though, for the first time, they were receiving "official" populations two years in a row (without two census surveys), the communes were few to react when reporting these numbers.

D. The cost of the device

46. A lengthy work has been necessary to implement this scheme: discussion with the tax administration, data evaluation, analysis on special cases, tests on data sets. Contributors were senior methodologists at the Directorate General of INSEE, and officials in charge of creating the survey and dialogue with the municipalities in the regional offices. We can estimate the statistical cost of the project to the equivalent of four years / person, good level methodologists. Must be added the development of tools able to accommodate sources and treat them, about a year of computer processor.

47. Although the operation is in its third completion in 2010, a full time statistician is still necessary to manage the device, plus the equivalent of another person spread over regional teams, computer or statistical support. In the following years, better pooling of the different source uses at INSEE (HT is also using in other statistical operations) could reduce this cost.
