

People with multiple residence and double counts in the French rotating census

Laurent Toulemon¹

Proposal for the 2016 European Population Conference
Theme 11: data and methods

Extended abstract

Introduction

Since 2004, the French institute of statistics (*Institut de la statistique et des études économiques*, INSEE) has organized a new type of population census, based on annual census surveys (Godinot 2005). The annual census surveys use a directory of addresses which is updated, leading to a likely lower proportion of omissions than classical general population censuses. On the other hand, conducting annual census surveys with a sampling rate of 14% may lead to increase of double counts, because most double counts are “invisible” for the individuals, if one only of their two usual residences is included in a survey.

The INSEE has published many papers on the processes (data collection, data coding) and disseminated many technical papers on its website (including estimates of the variance due to the sampling procedures) (Insee, 2015) but no empirical paper has been published on omissions and double counts.

The aim of the paper is twofold. First, to present, for the first time, an estimate of double counts within this new census. This is possible hope to the Demographic panel, a subsample which size has been set to 4% of the whole population, and which includes all census forms collected, including cases when two forms are filled the same year for the same individual. Second, some information is known on people having more than one usual residence, from another dataset, based on pooling many household surveys. These surveys use the same household grid, which includes questions on the existence of another usual residence. Comparing, on a macro basis, the proportion of people with multiple residence and the double counts will allow estimating the risk of double counts in case of multiple residence, for specific subgroups of the population.

Data and methods

The French census is now based on yearly census surveys

Since 2006, the census figures for each year N are estimated by merging five annual estimates in years N-2 to N+2. Each annual census survey is conducted on one-fifth of small municipalities (less than 10,000 inhabitants) and on 8% (1/12.5) of the population of large cities (10,000 or more)². As half the population is living in a municipality with less than

¹ French Institute for Demographic Studies, *Institut national d'études démographiques* (INED), Paris, France. toulemon@ined.fr

² In large cities, “large addresses” (grouping 60 dwellings or more, and cumulating no more than 10% of the population of a municipality) and “new addresses” (new dwellings appearing each year) are included with a probability of one-fifth.

10,000 inhabitants, the sampling rate is around 14%. Merging 5 years leads to a sampling rate of 70% (100% or 40% depending on the city size), data being adjusted to produce estimates at 1-1-N (See Insee, 2015 for more details).

The rules used to define the “main residence” of inhabitants are very complex and likely to lead to double counts, in case of multiple residence.

The French demographic panel, EDP

The INSEE is conducting, since 1968, a demographic panel (*Échantillon démographique permanent*, EDP) merging census data (based on individual forms and household information) and civil registration data (on births, deaths and marriages) for a one-percent sample. This sample has been enlarged to 4% since 2009, and the file contains no less than 390,000 individual census forms each year (Jugnot, 2014). When two forms are related to the same individual, both are kept, allowing an estimate of double counts (see below).

In march 2016, some information coming from tax files will be added to the EDP, allowing to get information on individual family situations and household structure and, if the merging procedure is accurate enough, to estimate omissions.

The pooled household grids

Since 2004, the INSEE is using a core set of questions at the beginning of most household surveys, describing the households in the dwelling (people sharing a common budget), and the list of households members. This list includes all people who live in the dwelling on a “usual basis”, even if they also have another “usual residence”, as far as they spend at least 30 days per year in the dwelling. This allows identifying people with more than one “usual residence”. Assumptions on the likelihood of double counts in these surveys (Toulemon and Pennec, 2010) led to the global estimates of 10% of households hosting at least one member having another usual residence, 7% of respondents having more than one multiple residence, and 4% of inhabitants having more than one usual residence (Toulemon and Denoyelle, 2012).

Estimating double counts at census from the EDP

From the EDP, the proportion of individuals included twice in a yearly census survey has been estimated in an unpublished INSEE note (Mambetov 2014). The number of identified individuals with two forms is around 1,000 each year, among 390,000 EDP members included in each yearly wave since 2009. The proportion of respondents included twice in the same wave, in two different dwellings, with a positive weight, is estimated as 0.24%.

The note did not propose any estimate of the number of double counts in the census that could be deduced from this figure. Considering a person having two dwellings, the probability that both dwellings are included in the census the same year depends on the situation of these dwellings. Excluding cases where both forms are filled by mistake in the same dwelling, five cases can be distinguished: 1. $P = 1/5$ if both dwellings are in the same “small municipality” (thus included together in the same wave); 2. $P = 1/5 * 1/5$ if the dwellings are in two different small municipalities; 3. $P = 1/5 * 1/12.5$ if one dwelling is located in a small city, the other in a large town; 4. $P = 1/12.5$ if both dwellings belong to the same address in a large city; 5. $P = 1/12.5 * 1/12.5$ if both addresses are in the same large city or in different large cities. In case of multiple residence, the probability that both dwellings belong to the same yearly sample thus varies from 1 to $1/12.5$, and the weight of people

with two residences varies from 5 to 156. The assumption of independence of the inclusion probabilities of the two dwellings must be checked carefully, in order to correctly weight individuals with two census forms the same year.

Estimating the probability of double count in case of multiple residence

Using people living in more than one usual residence (estimated from household surveys) as the group at risk, estimates of people counted twice in the census (from EDP) can be used to estimate probabilities of double counting among different subgroups of “commuters between households”, defined by age, sex, place of birth, enrolment in education and working situation, marital status, etc. In addition to these individual features, some categories could be used based on household information: family situation (living as a couple or not, living with parents or children, place of residence, etc.) but this information is related to one dwelling, making it difficult to use for people with more than one residence (see eg. Toulemon and Pennec, 2011, on people living alone in one or all of their dwellings).

First results

Around 4% inhabitants have more than one usual residence

Results from surveys show that multiple residence is more frequent among young adults, when they are still living in the parental home but also live in another dwelling (especially for their studies). Multiple residence is also frequent among children with separated parents, adults working in another city than their family home, and older adults around the retirement age who are spending part of the year in their secondary home (which is becoming their usual residence). The estimates vary from one survey to the next, despite the use of the same core questions by the same trained interviewers (Toulemon and Pennec, 2010), but an overall estimate of 4% “commuters between households” can be proposed.

Double counts at census could range between 1.4 and 1.9 %

The proportion of inhabitants actually counted twice in the census is estimated as 0.24% from the EDP. This includes only double counts where both dwellings are included in the same yearly wave of census. In order to transform this proportion of “concrete double counts” to “double counts at census”. If both dwellings were independent and the “commuters between households” at random in the dwellings, the multiplier would be $100/14 = 7.1$, leading to an estimate of 1.7% doubles counts in the French census. Estimates between 6 and 8 for the multiplier are likely, and 4 and 10 are possible, so that the proportion of double counts lies between 1.4% and 1.9% (between 1.0% and 2.4% at most).

Discussion and conclusion

The estimate of 1.7% is consistent with a proportion of 4% people having more than one usual residence, among whom 40% would be subject to double counts in the census. Double counts are very likely in household surveys for people with more than one usual residence, because the instructions for the fieldwork are very inclusive. In the census, on the contrary, the “main residence” rules applies, and inhabitants are supposed to fill one and only one form if they live in more than one “usual residence” (even if the rules are difficult to understand and to follow). The likelihood of double count at census depends on many factors. First, if one dwelling is unoccupied during the census days (between mid-January and mid-February), a double count is very unlikely to occur, as this dwelling will probably

considered as a “secondary home”, and their inhabitants thus excluded from the census. On the contrary, a child commuting on a weekly basis between both parental homes after a parental disruption is very likely to be counted twice. Similarly, young adults are likely to be included in the census in their parental home, even if they also live elsewhere for their studies (and should be included in the census only in their “studies-related” dwelling. Previous studies on double counts in the 1990 General population census led to estimate of 0.9% double counts, but the risk is inflated by the fact that most double counts are invisible, because both dwellings are most often not included in the same yearly wave (Desplanques 2008).

The EDP includes information on people included twice in the census from their both places of residence, allowing to accurately describing their family and social situation in both places. The EDP will be enriched with administrative data based on income tax declarations, leading to an extremely rich dataset; if data allows doing so, an estimate of omission probability in the census could be produced with comparing census figures and tax data, for the members of the Demographic panel. These data are expected to be released in March 2016, so that no commitment can be made now on their use. As tax data include, on an individual basis, age, sex, and family situation, omission probability estimates could be produced for the same groups of population than double counts. If the quality of the merging process is sufficient, a new avenue of research could be open on census quality.

These new analyses first imply that a consensus on double counts is reached, and the current paper will be a step towards more collaboration with INSEE on the “hot” topic of census accuracy in France.

References

Desplanques G., 2008, “Strengths and Uncertainties of the French Annual Census Surveys”, *Population-E*, vol. 63, n° 3, p. 415-439.

Godinot A., 2005, Pour comprendre le recensement de la population. *Insee méthodes*, <http://www.insee.fr/fr/publications-et-services/sommaire.asp?codesage=imeths01>

Insee, 2015, Documentation sur les résultats du recensement. <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/documentation-aide.htm>. Last update: 15 October 2015.

Jugnot S., 2014, La constitution de l'échantillon démographique permanent de 1968 à 2012, Documents de travail de l'Insee, n° F1406, http://www.insee.fr/en/themes/document.asp?reg_id=0&ref_id=F1406

Toulemon L., Denoyelle T., 2012, « La définition des ménages dans les enquêtes françaises : comment tenir compte des multi-residences ? », *Actes des journées de méthodologie statistique*, 16 pages.

Toulemon L., Pennec S., 2010, “Multi-residence in France and Australia: Why count them? What is at stake? Double counting and actual family situations”, *Demographic Research* 23(1): 1-40.

Toulemon L., Pennec S., 2011, “How many people live alone in France?”, *Population & Societies*, 484, 4 pages.