AN ALTERNATIVE TO TRADITIONAL CENSUS TAKING: PLANS FOR FRANCE

Jean Dumais, Sylvie Eghbal, Michel Isnard, Michel Jacod, François Vinot,
Institut National de Statistique et d'Études Économiques (France)

Introduction

Over the last decade, reductions in public spendings have been observed in many Western countries. In a number of countries, it has thus become increasingly difficult for national statistical offices to justify and obtain sufficient funding for a complete enumeration of the population. Moreover, in a growing number of European countries, the masses of administrative information in the custody of government agencies has made the need for a complete decennial[1] enumeration - technically at least- less acceptable by the tax payers. In view of these environmental changes, France has embarked on a complete overhaul of its general Census.

Use of administrative data, combined with sampling and small area estimation techniques will allow reductions in response burden, balanced costs over time, while maintaining legal and statistical legitimacy.

Legal Requirements

Every year, a « legal » population count has to be published. This number serves as a basis for the distribution of public monies to the « *communes* » or municipalities[2]. Traditionaly, the « legal » population would be the count at the latest general census[3].

Given that INSEE has chosen to proceed to an annual (partial) census, any additional changes to the way the census is conducted must account for the requirement of an annual « legal » population size.

Contrary to United States or Canada, there is no constitutional or legal requirement to take a census in France even though a number of acts make specific reference to a general census of the population. In fact, "census is not even a matter of law ; it is conducted (or not, when budgetary constraints are too strong) after an administrative act and parliament is only involved when voting the budget."[4] There has been a concensus to conduct a general census every 7 to 9 years, making sure not to interfere with election cycles.

---

[1] In France, the last 5 censuses were conducted in 1968,1975, 1982, 1990, and 1999; though a census was planned for 1997, it was postponed to 1999 in view of Maastricht agreements .

[2] Remembering that Paris comprises 20 *communes*, Marseille 16, and Lyon 9. There are 36,685 *communes*, half of them with fewer than 400 inhabitants.

[3] Occasionally, a municipality may ask to be recounted if it feels its growth has made its legal size obsolete.

[4] Deville, J. C., Jacod, M., *Replacing the traditional French census by a large scale continuous population survey*

Statistical Requirements

In addition to the « legal » population, there is a « statistical » population. It is what we expect it to be : the latest census count adjusted, through vital and other administrative records, for births, deaths and migration.

Whatever statistical activity the census becomes, the need for a « statistical » population remains and must be met, for each municipality. We have set as a goal to make any part of the country - even sub-municipal - publishable (obviously within the limits of confidentiality) as it would with any traditional census; recourse to administrative data and modelling will allow this requirement to be met annualy rather than at "the end" of a 5-year collection cycle.

Sampling Strategy and Administrative Sources

The pivotal unit is the *commune*: smaller *communes* (less than 10,000 people) would be sampled at the (average) rate of 1/5, and then completely enumerated; all the larger *communes* would be covered every year, and only a fraction (about 1/12) of the dwellings would be contacted. As the smaller *communes* comprise about one half of the population, the overall sampling annual sampling fraction would be about 1/8.

For the domain of the smaller *communes*, we would create 5 rotation groups of *communes* to be surveyed in turn. Those annual rotation groups would be created within each of the 22 regions. Using 1999 census data, each group would be created so as to best represent each region socio-demographic make-up, thus minimizing the year-to-year variation. Within each region and each rotation group, a complete enumeration would take place as is done in a traditional census operation.

For the domain of the larger *communes*, we plan to use a "building register". The "building register" is a list of buildings (residential, institutional or commercial) uniquely identified and located so as to create a set of digitized maps. The "building register" will be populated with the results of the 1999 General Census of the Population; hence each building will be further described by a number of sociodemographic statistics. The register will be continuously updated using building / demolition permits, utilities billings, information from local authorities and direct observation. Thus, the « building register » can serve as a sampling frame in the larger *communes*. The sampling plan considered for each larger *commune* is a stratified 2-stage sample of dwellings. Firstly, the buildings in each *commune* are stratified in "IRIS2000"[5] ; in each "IRIS2000", five rotation groups are created as seen above for the smaller *communes* and the groups are selected in turn for the annual survey. Secondly, in the annual rotation group, the listed buildings will be visited and a complete list of dwellings will be drafted; from this list, a random sample of about 40% of the dwellings will be drawn.

Data Collection and Processing

Almost a natural consequence of the continuous enumeration, the data will be collected on the longest time span possible: September Y-1 to June Y; the collected data will later be corrected to the reference date of 1st January Y. This longer collection period will allow more economical training and development of census takers and increased data quality. As well, the continuing improvement of techniques and tools can be sooner implemented thus avoiding costly and disruptive major data processing systems redesign. It should be noted that the collection campaign for a given *commune* is planned to last about a month, the current duration of the census data collection.

Estimation and Dissemination

Annual estimates for the larger *communes* should be straightforward expansion-type estimates. Annual estimates for smaller *communes* will be possible for those of them subjected to the annual CCP.

The current dissemination plans are to publish by December 31, Y, the national results of the census taken in year Y and the results for some large segments of the land (e.g. metropolitan areas of more than 1 million inhabitants). As well, the results for the smaller *communes* enumerated during Y would be made available. Moreover, the detailed results of Y-2 [6] would be made available; these detailed results would come from a census or sample survey (smaller or larger *commune* respectively) or from synthetic estimation. A synthetic estimate would make use of the relationships between observed and administrative data for a given area at a given time point. At this time, we plan to use administrative data at a very high level of geographical detail (building, city block, enumeration area) that would tell us something of the people (age-sex distributions from health care files "HC" in what follows) or their dwellings (number of dwellings from building register, utilities billing lists or dwelling tax roll "DT" in what follows).

---

[5] IRIS = "*îlots regroupés selon des indicateurs statistiques*" or city blocks regrouped along statistical indicators; each grouping is connected, rather homogeneous and comprises about 2,000 people.

[6] The 2-year lag is what is currently expected to be the time needed to receive and process the administrative data.

ROTATION GROUP					Dissemination Reference Date		Current Year
	Y-6	Y-5	Y-4	Y-3	Y-2	Y-1	Y
GROUP 1					C	S	S
GROUP 2	C				S	C	S
GROUP 3		C			S		C
GROUP 4			C		S		S
GROUP 5				C	S		S

C = census taken
S = synthetic estimate

For a given small *commune* surveyed in Y-5 and Y (see diagram below), a number of "person" variables (age, sex, occupation, ...) and "household" variables (household size, size of dwelling, ...) are measured both times. From the administrative sources, we have complementary information at a fine level of detail, and a distorsion between what was seen and what is logged on administrative files can be measured on similar aggregates (e.g., building or city block). Hence, for the observed aggregates, a correction factor can be computed so that administrative counts add up to the census estimates. For example, for year Y-4, $P_{A-4} = HC_{A-4} / HC_{A-5}$ and $L_{A-4} = DT_{A-4} / DT_{A-5}$ are factors to be applied to Y-5 "people" and "household" records respectively. Synthetic estimation for Y-3 to Y-1 can be made in a similar fashion.

Moreover, synthetic estimation for Y-4 to Y-1 could benefit from the information obtained during the most recent census in Y; one could compute the adjustment factors with reference to the latest census estimates, *viz.* $P_{A-2} = HC_{A-2} / HC_A$ and $L_{A-2} = DT_{A-2} / DT_A$, and backcast the intercensal period.

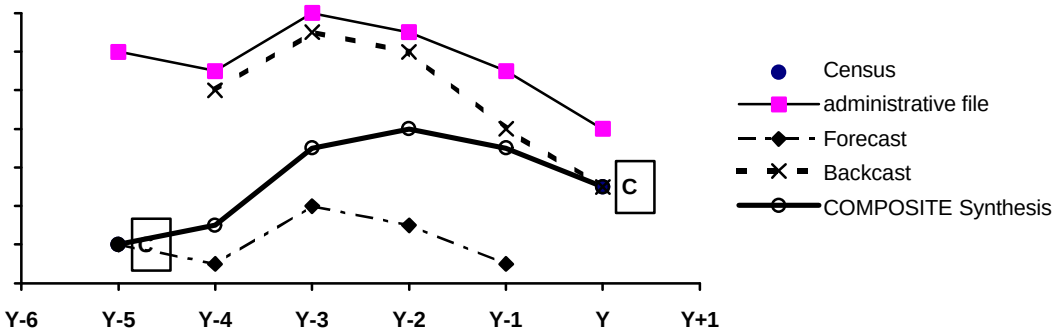
ROTATION GROUP					Dissemination Reference Year		Current Year
	Y-6	Y-5	Y-4	Y-3	Y-2	Y-1	Y
GROUP 1					C	→S	→S
GROUP 2	C				→S ←S	C	→S
GROUP 3		C	→S ←S	→S ←S	→S ←S	→S ←S	C
GROUP 4			C		→S		→S
GROUP 5				C	→S		→S

C = census taken
→S = synthetic estimate be forecasting
←S = synthetic estimate by backcasting

Chances are that the two series of synthetic estimates will not coincide. Yet, it is highly desirable to use and publish a single series of estimates for any time point and any geographic zone. Our proposal is to blend the series of synthetic estimates into a "composite" series anchored at both ends to the census counts. The following linear combinations would be used:

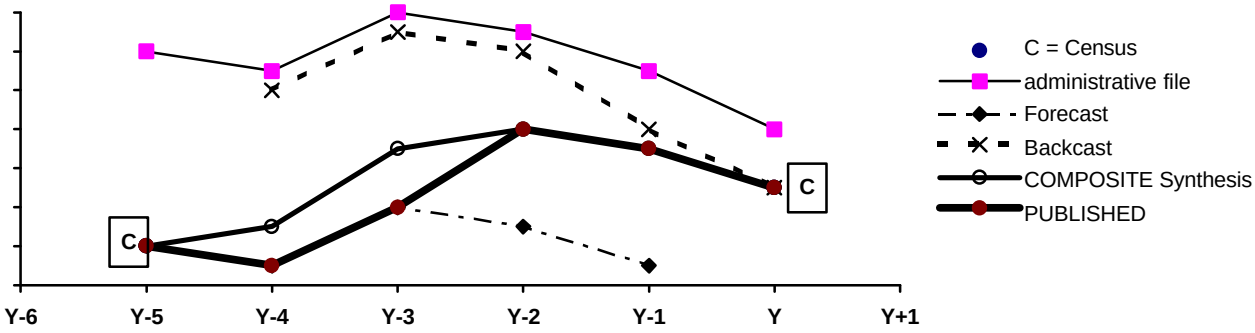
$$S_{A-4} = 0{,}8 \times Forecast_{A-4} + 0{,}2 \times Backcast_{A-4}$$
$$S_{A-3} = 0{,}6 \times Forecast_{A-3} + 0{,}4 \times Backcast_{A-3}$$
$$S_{A-2} = 0{,}4 \times Forecast_{A-2} + 0{,}6 \times Backcast_{A-2}$$
$$S_{A-1} = 0{,}2 \times Forecast_{A-1} + 0{,}8 \times Backcast_{A-1}$$

The graph below illustrates how these series could behave, given in this example that the administrative sources differ from the Census by as much as 10% at the beginning of the cycle and by as little as 1% five years later.



In summary, for a given *commune* users would be offered the following data series:

timeline	Y-5	Y-4	Y-3	Y-2	Y-1	Y	Y+1
published data	- Census count for Y-5 - composite for Y-7	Composite for Y-6	Count for Y-5 - revisions for Y-9 to Y-6	Forecast for Y-4	Forecast for Y-3	- Census count for Y - composite for Y-2	Composite for Y-1 - revisions for Y-4 to Y-1



The Impact on Data Series

The impact of the CCP on socio-demographic data series is not ignorable: age-sex, marital status, labour force status distributions, mobility etc. have now to be determined from sample surveys. Disturbances in data series due to annual sample rotation are to be minimized.

Work in Progress

Work is on-going on optimal stratification and sample allocation. A rule must be devised to allow the promotion of *communes* to, or demotion from, the « large *communes* » stratum.

Exploration and assessment of administrative sources has started; obviously this is a crucial element of the whole programme, given the reliance of the CCP on administrative sources.

Small area and synthetic estimation are explored in parallel. Close attention is paid to data series benchmarking, annual calibration and variance estimation.

Last but not least, discussions with users on the redesign of the census have started on three broad themes: infra-municipal statistics (public transportation, location of retail trade, services, etc), supra-municipal statistics (population dynamics, employment basins, etc), and socio-demographics (family, household, dwellings, etc). These users-Insee meetings are to be held this Summer ; discussion groups will also take place during Summer and Fall 1999 with the mayors and local authorities. The outcome of these discussions sould be brought to the *Assemblée Nationale* as a bill in early 2000. Field work is planned to start in Autumn 2001.

Conclusion

The CCP breaks away from the long-established tradition of actually counting people once every 7 to 9 years; the CCP is a census as it can produce local estimates of population (size and structure); it is innovative as the collection is continuous and census estimates continuously updated.