TWINNING CONTRACT

CRIS 2015/370-467

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

MISSION REPORT

on

Component D
Component D: Methodological and geo-spatial tools for improving the quality and efficiency of field surveys

Activity D.7
Integration of sampling, workload allocation, management and monitoring in a multi-field survey approach

Implemented by

- *Mr. Georges Bourdallé, Program Office, INSEE. georges.bourdalle@insee.fr*
- *Mr. Pierre-Arnaud Pendoli, Statistician responsible of population census sampling design INSEE. pierre-arnaud.pendoli@insee.fr*

Jerusalem

05 – 08 February 2018

Version: Final

Institut national de la statistique et des études économiques

Insee

Mesurer pour comprendre

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

Authors' names, addresses, e-mails

Mr. Georges Bourdallé
Program Officer
National Institute of Statistics and Economic Studies France (INSEE)
Email: georges.bourdalle@insee.fr

Mr. Pierre-Arnaud Pendoli
Statistician, responsible of population census sampling design
National Institute of Statistics and Economic Studies France (INSEE)
Email: pierre-arnaud.pendoli@insee.fr

Table of Contents

Annexes produced for the mission (external to the report):
Annex D7 - 1 Terms of Reference D7
Annex D7 - 2 Meeting program for D7
Annex D7 - 3 Persons met at activity D7
Annex D7 - 4 Introduction to the Survey Department and activity D7
Annex D7 - 5 Presenatation of MSS
Annex D7 - 6 Overview of household surveys in France
Annex D7 - 7 Introduction to CAPI in France
Annex D7 - 8 Sampling methodology for French social surveys

List of Abbreviations

BC	Beneficiary country
CAPI	Computer Assisted Personal Interviewing
EU	European Union
Fideli	French acronym meaning "demographic file ofhousings and individuals
IAA	Interviewer Action Area
ICBS	Israeli Central Bureau of Statistics
INSEE	National Institute of Statistics and Economic Studies of France
IT	Information technology
MS	Member State (of the EU)
NAUTILE	Nouvelle Application Utilisée pour le Tirage des Individus et des Logements dans les Enquêtes (Nautile : French acronym meaning "New application used for housings and individuals surveys sampling")
OCTOPUSSE	Organisation Coordonnée de Tirages Optimisés Pour une Utilisation Statistique des Échantillons
OPALE	Outil de planification des activités liées aux enquêtes
PSU	The Primary Sample Units
SSU	Secondary Sample Unit
ToR	Terms of Reference

Executive Summary

During the four day mission, the MS experts from National Institute of Statistics and Economic Studies of France (INSEE) were introduced to the vision of an integrated workload allocation, management and monitoring in a multi-field survey approach for household surveys at The Israeli Central Bureau of Statistics (ICBS). The vision is to integrate all stages and collections methods into one management and monitoring system to be used by all actors involved in field surveys. Thus the principal goal is the integration of sampling, workload allocation, management and monitoring into a multi-field survey approach. To do this, it is necessary to distinguish between two parts: Firstly, organization management of the enumerators and the surveys and planning the workload. Secondly, the sampling methodology. The ICBS has a long term vision of integrating all this into what has now been named as the MSS or Multiple Survey System.

The ICBS also presented the current situation regarding planning, management and monitoring of field survey and the MS experts presented the experiences of INSEE on planning and monitoring social surveys.

Survey management:
In brief some differences between the INSEE practice and the ICBS practice were encountered but these are not incompatible with a multi-field survey approach (see section 2). However, some six recommendations for optimizations were proposed by the INSEE experts. In brief they were (for details please consult section 3):
* *To increase the coordination between services or actors.*
* *To create a new coordinator function responsible for the overall coordination of field surveys*
* *To optimize the design of questionnaires*
* *To standardize the computer collection applications of the enumerators and managers*
* *To develop a tool to estimate the workload of the enumerators (including training, vacation…)*
* *To improve the planning*

Sampling methodology:
ICBS and INSEE both use self-weighting two-stages sampling design. Indeed this sampling design guarantees the equality of the sampled unit's weights. At the first stage of the sampling INSEE draw the primary sample units (PSU), previously built from a geographical point of view, with inclusion probabilities proportional to the size. At the second stage of the sampling, INSEE draw the housings or the individuals in the PSU. Except for the PSU whose inclusion probabilities are equal to 1, the allocations between PSU are equal.

However, MS experts noticed that there are differences in the requirements of INSEE compared to the requirements of the ICBS. In France, the PSU are drawn once to draw almost all the surveys samples during a decade. For instance, INSEE's methodologists are currently testing several methods to draw the unique PSUs of the New French Master Sample that will be used to draw the samples and will be used from2020-2030. This specificity comes from the fact that INSEE's enumerators have to be hired for a decade to collect the samples in fixed areas. In Israel, the ICBS draws a specific annual primary units sample for each survey.

In addition, the MS experts noticed specificities concerning the management of the samples that is integrated in a computing application at INSEE.

As an outcome of this mission the ICBS and INSEE experts agreed on four recommendations for sampling methodology outlined in the third part of the report. In brief they were (for details please consult section 3):
* *To build the primary units or the interviewers action areas by using the Travelling Salesman Problem*
* *To use the spatially balanced sampling design to draw the PSU*
* *To use negative coordination methods for the first stage or for the second stage of the sampling*
* *To create a computing application to manage the sampling.*

1. General Comments

This mission report was prepared as part of the Twinning Project"*Support to the Israeli Central Bureau of Statistics (ICBS) in Enhancing the Quality of Official Statistics"*. It was the seventh activity within Component D: Methodological and geo-spatial tools for improving the quality and efficiency of field surveys and the mission was devoted to *Workload allocation, management and monitoring of multi-field social surveys*

The MS experts would like to express their thanks to all officials and individuals met for their kind support and valuable information which they received during the stay in Israel and which highly facilitated the work of the experts.

The views and observations stated in this report are those of the MS experts and do not necessarily represent the views of EU, ICBS or INSEE.

During the mission, the ICBS experts and the MS experts presented and discussed the benefit and challenges in implementation of multi-field survey system. The INSEE has extensive experience with applying a multi-field survey approach. The MS experts presented the responsibilities and task for all actors involved in the collection process as well as tools developed to support the data collection. They presented CAPI[1] for computerized collection of surveys, OCTOPUSSE[2] and NAUTILE[3] for the sampling and OPALE[4] for estimating the workload of the enumerators in a multi-field approach. The MS experts pointed out the importance of including time for training and vacation days in the multi-field survey approach in order to properly estimate the workload of the individual enumerators.

The ICBS aim is to integrate all the tasks of the data collection in their *Multiple Survey System* (MSS). For historical reasons, INSEE do not have a similar system. The MS experts believe that combining all the tasks in one system such as outlined for the MSS system is a good and visionary approach applied by the ICBS.

Concerning sampling methodology, the MS experts noticed that the general principles are the same at the ICBS and at the INSEE. Both organizations use self-weighting two-stages sampling design. The equality of the sampling weights is to prevent the estimations from having too big an influence of an error of collection for some sampled units. The main difference between the INSEE and the ICBS sampling design occurs at the first sampling stage.

- In Israel, the sampling is specific to each survey from the first to the second stage: Each year new PSU are drawn and the secondary sampled units (SSU, inhabitants or dwellings) are drawn from the annual PSU. Before the sampling, ICBS partitions the Israeli territory into primary units by considering each locality as a PSU.

- In France, the second stage of the sampling is specific for each survey. But the PSU are drawn once for all the surveys for the period of a decade. Indeed, INSEE manages a fixed network of interviewers who are hired for a period of ten years in order to collect all the social surveys in fixed areas. For instance, as soon as the Sampling Division at INSEE draw the PSU for the new French Master Sample, the enumerators will be hired in these PSUs for the decade 2020-2030 and all the samples will be drawn in these areas for these ten years. Before the sampling, INSEE partitions the French territory by aggregating the localities to build its primary units. In fact, the purpose of the PSU building is to minimize the geographical extent of the PSU under constraints of sizes (in terms of number of main residences).

Besides that the MS experts also noticed that the two core sources for sampling– the Central Population Register and the Dwelling Register – used by the ICBS to draw the social surveys samples check the expected quality requirements for a sampling frame.

---

[1]Computer Assisted Personal Interviewing
[2]Organisation Coordonnée de Tirages Optimisés Pour une Utilisation Statistique des Échantillons
[3]Nouvelle Application Utilisée pour le Tirage des Individus et des Logements dans les Enquêtes
[4]Outil de planification des activités liées aux enquêtes

- Concerning the units (dwellings or inhabitants), each sampling frame is an exhaustive basis updated regularly (annually for the Dwelling Register and three times a year for the Central Population Register).

- Concerning the variables, each sampling frame has advantages and disadvantages:
  o The Central Population Register gathers a lot of socio-demographic variables but geographical variables are poor. However, by using an address geo-coding application, exact (x, y) coordinates of nearly 75% of the addresses in the register can be found.
  o On the contrary, in the Dwelling Register ICBS doesn't have any information about the households corresponding to the dwellings but several geographical variables are available in this sampling frame.

Lastly, at the second stage of the sampling for both ICBS and INSEE use a systematic sample.

2. Assessment and Results

All the foreseen activities were carried out following the plans in the ToR; cf. Terms of Reference (Annex B3 - 1). The conclusions are described in the following section.

ICBS presented:
- An introduction to the Survey Department
- Presentation of the use case
- Presentation of the MSS System
- The sampling frame used in Israel for field surveys today and challenges foreseen in moving to a multi-field approach with Interviewer Action Area (IAA)

The MS experts presented:
- Introduction to field surveys in France
- Organization and management of field surveys in France
- Sampling, creating workloads and allocation interviewers for a field work in France
- The CAPI system

There are some differences between the practices of INSEE and the ICBS. However, these differences are not regarded as incompatible with the multi-field survey approach. The practices are mainly a result of historical decisions taken in each statistical institution

Organization and the management of the collection
Differences between the INSEE and the ICBS include:
- The status of the enumerators. At the ICBS, the enumerators have a contract with the ICBS and they are paid according to their working hours. At INSEE, since 2013, the enumerators are an integral part of the staff at INSEE. This difference is important and influences the workload estimation of the enumerators. When the enumerators are part of the office staff they seem to stay for a longer period of time and thus, it seems simpler to plan the workload. However, the question of the status of the enumerators of the statistical bureau is more complicated than the question of the estimation of workload.
- Concerning the collection of household surveys, ICBS use two applications, one computer application for the enumerators and a computer application for the managers.
- Concerning the questionnaire: ICBS develop the questionnaires in different softwares and organize each survey in a different way. In addition, each survey has a specific different questionnaire.
- Concerning the planning of the surveys. INSEE plan the survey's program three years in advance. This allows a good view of the adjustment needed in the regional office and enumerators and to estimate the busy times.

- Concerning the regional manager. At the ICBS the regional manager is in charge of a portfolio of enumerators and at INSEE the regional manager is in charge of a portfolio of surveys. INSEE prefer to have regional managers specialized in certain surveys (one or more).
- ICBS has two different applications of sampling, one for the LFS survey and one for the household survey.

Sampling
During the general presentation of the sampling methodology used at INSEE (Tue 5 Feb) and the parallel session later (Wed 6 Feb), the principles of the French Master Sample were presented and discussed.

- **Concerning the building of the primary units**. In Israel, each locality is considered as a primary unit, independently of its size. As the PSU inclusion probabilities are set proportionally to their sizes, it means that the smallest localities have nearly no chance to be drawn in the first stage of the sampling. But we have to notice that the huge majority of the Israeli population lives in urban cities and not in rural areas. However, similar to INSEE, it could be a possibility to gather the localities in order to build PSU with a minimal size (in terms of housings or individuals). In France, the Sampling Division used the algorithm of the Travelling Salesman Problem in each *département*[5] to build the PSU. The underlying idea of the use of this method is to minimize the extent of the PSU in order to limit the interviewers' travel time and thus to limit the costs of a face-to-face survey. This method is easily computable in R thanks to the TSP package. The algorithm needs the two-dimension distance matrix by road between all cities. For each starting city, we get a path that is the solution to the algorithm. From a certain starting point we travel the cities located on the path – which is the solution – until reaching the minimal size of a PSU. After we build the first PSU, we continue to use the path to build the following PSU.

- **Concerning the first stage of the sampling**. Several sampling designs are possible to draw the PSU while respecting their inclusion probabilities. At INSEE, for the new French Master Sample, the Sampling Division is currently using the spatially balanced sampling method in order to geographically disperse the PSU sample when a strong spatial autocorrelation for some variables is noticed (like the median income of the households or the socio-professional structure of the population). This sampling design, proposed by Grafström & Tillé (2013), was developed with the package "BalancedSampling" in R. Compared to a simple balanced sampling design, it consists in adding the (x, y) coordinates of the biggest city as a balancing variable of the PSU. Simulations of this sampling design show gains in terms of accuracy and geographical dispersion, compared to a simple balanced sampling design.

- **Concerning the survey burden for the households**. By drawing independently several samples in the same sampling frame, a unit has a probability to be drawn in several samples. For Business surveys, INSEE uses negative coordination methods in order to spread out the survey burden on businesses. In practice, the purpose of negative coordination is to take into account the previous surveys when drawing a new sample in order to minimize the overlap between samples by fostering the selection of a unit that has not already been selected in recent surveys while preserving the unbiasedness of the samples. This method could be used by INSEE at the second stage of the sampling to draw the housings/the individuals. At ICBS, this method could also be used at the first stage of the sampling in order to decrease the overlap of the annual surveys PSU. However we have to take into account the fact that this method would modify the sampled unit's weights.

- **Concerning the computing management of the samples**. INSEE's social survey samples are not drawn thanks to separate SAS or R programs. Since 2009 a computing application called "Octopusse", developed in Java, has been used to draw the samples in the annual census surveys. Starting 2020, Octopusse will be replaced by a new application, whose name will be

[5] In France, the regions are divided in administrative areas called "département(s)".

Nautile in order to draw the samples in the future new sampling frame called Fidéli[6] that gathers several tax files (e.g. housing taxes etc.). A computing sampling application is very useful to manage the disjunction of the samples and the setting of the sampling designs parameters.

3. Conclusions and Recommendations

The multi-field survey approach requires optimization of all the tasks in the data collection. There are two types of recommendations. The recommendations for multi-field survey management and the recommendations for sampling methodology in a multi-field survey approach.

Concerning the multi-field survey management, six recommendations are proposed by the MS experts:
- To increase the coordination between services or actors. It is necessary to ensure that all the actors in a survey (methodology, survey designer, manager, IT teams) have the same level of information and understanding.

- To create a new function for the coordinators. The MS experts presented the CPOS (Project manager in Organization Statistics) in France. The multi-field survey approach is like an orchestra, you need a conductor.

- To optimize the design of questionnaires to weigh in development time and include all the questionnaires in the MSS system
    o The MS experts propose to build a questionnaire in two parts. A common part for the household variables such as the composition of the household, demographic variables and a second part which is survey specific.

- To standardize the computer collection applications of the enumerators and managers
    o Build the same computer application system for enumerators managers.
    o Build one computer collection application for all surveys including LFS.

- To develop a tool to estimate the workload of the enumerators (including training, vacation…)
    o The MS experts propose to build a system such as OPALE which can be included in the MSS system

- To improve the planning
    o The MS experts propose to plan the surveys program and the time period of the collection for 2 or 3 years in advance. In a multi-field survey approach it is important to estimate and to anticipate the possible difficulties of an early collection.

Concerning sampling methodology, the MS experts presented four recommendations to the ICBS:
- To build the primary units or the interviewers action areas (at INSEE, PSU and IAA overlap for the small localities), the ICBS can use the Travelling Salesman Problem whose algorithm is easily computable in R thanks to the TSP package.

- To use the spatially balanced sampling design to draw the PSU in order to improve the quality of the estimations in terms of accuracy and geographical dispersion.

- To use negative coordination methods for the first stage or for the second stage of the sampling in order to spread out the survey burden among the housings/the individuals.

- To create a computing application to manage the sampling. Since 2009, INSEE has used an application called "Octopusse" to draw the samples in the last annual census surveys. Starting 2020 to 2030, INSEE will use a new application named "Nautile", which is still in development, to draw the samples in the new sampling frame called Fideli.

---

[6]French acronym meaning "demographic file of housings and individuals"

Support to the Israeli Central Bureau of Statistics in Improving the Quality of Official Statistics

Literature

Grafström, A., & Tillé, Y. (2013): Doubly balanced spatial sampling withspreading and restitution of auxiliary totals. *Environmetrics*