The Coordination of Business Survey Samples

Emmanuel Gros & Thomas Merly-Alpa

***Résumé* — The aim of this methodological note is to provide a brief description of the new methods being implemented at INSEE for the coordination, both positive and negative, of the samples drawn for business surveys.**

This note is largely based on the article by E.Gros in letter no. 73 within the ESS, dated July 2016 (internal INSEE document).

I. The Coordination of Samples – Principle and Procedure Implemented at INSEE

A. Principles and Objectives

The aim of coordinating samples is to take account, when drawing a given sample, of those used for previous surveys, generally with a view to reducing the statistical burden on businesses while maintaining the unbiased nature of the samples.

There are two separate types of coordination. La **Negative coordination** involves prioritising the selection of businesses that have not been selected for any recent surveys. It forms part of an effort to reduce the statistical burden on small enterprises, since large enterprises (beyond a certain threshold) are routinely questioned in the majority of surveys. Conversely, the aim of **positive coordination** is to maximise the overlap between coordinated samples, either for panellisation purposes or once again with a view to reducing the overall statistical burden, this time by reducing the size of survey questionnaires on similar topics (the responses to a survey can be used for another without being collected via that survey).

B. The Procedure Used at INSEE Since the End of 2013

The coordination method in use at INSEE [1] lows samples to be drawn in accordance with a coordinated **stratified simple random sampling plan** [1] and is based on two fundamental concepts : the permanent assignment of a random number to each unit within the business sphere and the use of coordination functions. More specifically, the coordinated sampling procedure works as follows :

— each unit is permanently assigned [2] a random [3] and invariable number, with which it remains associated for its entire existence ;

— when drawing a sample, the units are selected by drawing the units with the smallest permanent random numbers within each stratum, converted via a coordination function, determined with a view to ensuring that the selection of a sample satisfies the coordination objective being sought ;

— the coordination function is selected taking account of the cumulative burden function associated with past surveys, which is calculated on the basis of the coordination functions used during sampling for those surveys.

This coordination procedure offers many advantages as it allows :

— **any number of surveys to be coordinated** with one another. It is even possible to positively coordinate the sample with certain past surveys and to negatively coordinate it with others ;

— differentiated coordination to be managed in accordance with the surveys by applying a weighting to the sampling during the coordination process based on the duration of the questionnaire or association with a panel for example ;

— **samples belonging to different levels to be coordinated with one another** : groups, profiled businesses (EP), legal units (LU) and/or establishments. This multi-level coordination presupposes the establishment of unambiguous and lasting links between units belonging to different levels : each group has a single main PE, each PE has a single main LU and each LU has a single main establishment. The units linked in this way are assigned the same permanent random number, which then guarantees multi-level coordination. This coordination is therefore partial : where a LU has multiple establishments, for example, only the main establishment will be assigned the burden of its LU.

II. Performance of the Coordination Procedure

Thanks to simulation studies, it has been possible to quantify the effectiveness of this coordination procedure in terms of the distribution of the response burden across the various units within the population. The method consists of linking the drawing of twenty LU samples, with each sample then being negatively coordinated with the previous samples [2]. Table 1 ompares the distribution of the “survey burden” variable (defined as the number of samples to which a unit belongs) resulting from this coordinated sampling

1. This is the standard sampling plan used for business surveys at INSEE.

2. During the initialisation of the system for units already existing on that date, or during their creation for the other units.

3. Following a uniform distribution on [0 ; 1]

with that obtained by means of independent sampling.

Survey burden	Independent sampling	Coordinated sampling
0	3 981 423	3 952 718
1	391 840	445 402
2	30 494	9 084
3	3 670	606
4	374	9
5	18	0

Fig. 1. Survey burden for 20 independent or coordinated samples

As anticipated, there is a narrowing of the distribution around 1, i.e. spreading of the survey burden : the number of units surveyed more than once decreases significantly, as does the number of units that have not been surveyed, in favour of a very clear increase in the number of units selected for a single survey. Moreover, **this coordination method is proving to be more effective** [3] **than the method previously used at INSEE** [4], which only allowed surveys to be coordinated in pairs.

III. Constraints Inherent in the Coordination Procedure

The introduction of this global coordination procedure imposes limits with regard to the coordination of samples during the first year (or the first X years for rotating panels renewed X times) when they enter into the system. Indeed, **coordination can only take place with samples drawn using the new method** as it is necessary to use the permanent random number in order to draw the sample on the one hand and to have the coordination functions used previously to draw samples on the other hand in order to perform the coordination.

More specifically, this means that it is impossible to coordinate the drawing of the sample for a given survey with survey samples drawn before this coordination procedure was introduced at the end of 2013. In the case of rotating panels in particular, it is not possible to coordinate with past panels during the renewals performed during the first few years. Therefore, for any rotating panel renewed by halves/thirds/quarters/etc., the coordination with past panels can only be "complete" once the entire rotating sample has been renewed using the new coordination procedure.

n addition, this coordination procedure is **incompatible with the selection of stratified samples with the systematic drawing of sorted data from within the strata**, which was previously frequently used at INSEE in order to obtain, within each stratum, a distribution of units within the sample that is close to that observed in the sampling frame for the sorting criterion. Nevertheless, since the systematic drawing of samples is linked to implicit stratification with proportional allocations, it is possible to take account of the criterion formerly "controlled" by the systematic sampling in the coordinated sampling method by rendering this stratification explicit by means of an "overstratification" method [2].

IV. Examples of Application

A. Traditional Surveys

For the majority of INSEE and SSP surveys, negative coordination is ensured with the other surveys gathered during that same year in order to limit the survey burden. This forms part of the simplification of public policy [5], which aims to allow small businesses to only respond to one mandatory non-European survey each year. The surveys are often also coordinated with previous surveys covering similar topics or fields so as not to provoke survey burnout among the businesses.

B. Business Outlook Surveys

Within the scope of the monthly Business Outlook Surveys, the renewal of the samples from one year to the next is always a complex operation, as the stability of the indicators must be guaranteed. The positive coordination method ensures optimal overlap of the different annual samples while maintaining their bias-free nature and enabling negative coordination to be performed with other surveys in order to avoid a heavy collection burden, since monthly surveys are often quite burdensome for businesses.

However, in this case, it is not possible to fully control the number of units within a sample that are retained from one sample to the next or which units are surveyed again. The date on which units exit the panel can also vary.

At INSEE, surveys such as the EMAGSA (monthly survey of large-scale food retail activities) use this method with the aim of also fostering loyalty among the businesses and therefore facilitating contact with the units surveyed.

C. Rotating Panels

The coordination method described here can be used for rotating panels. Indeed, a rotating panel is a system by which a part (for example half or a fifth) of the sample is renewed each year while the other parts remain unchanged. This ensures measurable stability for the sample. Negative coordination with the sample on the date of entry into the system therefore makes it possible to ensure that the "outgoing" units are not reintegrated into the panel for a new cycle ; not only does this ensure that the collection burden is limited for the longitudinal units, it also ensures that the entire population of interest is more effectively

covered.

The ICT (Information and Communication Technologies) surveys and the ESA (Annual Sectoral Survey) performed up until 2016 make use of this method in order to eventually gather structural information for all French businesses while also ensuring that the results remain consistent from one year to the next.

D. More Atypical Cases

inally, in some cases, the coordination method is implemented in order to comply with constraints that may appear contradictory. Within the scope of an extension to the EACEI survey (annual survey on industrial energy consumption) aimed at disseminating accurate results at regional level, the question arose as to whether it should be coordinated with the Antipol (anti-pollution investments) survey, which covers the same field and the same topics. Indeed, the sample sizes were so large relative to the field that negative coordination between the EACEI survey, its extension and the Antipol survey would result in the same units being surveyed again. It was therefore necessary to conduct a specific study in order to determine the strata in which sampling was to be positively coordinated and in which it was to be performed separately.

REFERENCES

[1] Fabien Guggemos et Olivier Sautory, « Sampling Coordination of Business Surveys Conducted by INSEE », Proceedings of the Fourth International Conference of Establishment Surveys, Montréal, Canada, 11-14 juin 2012 (https://ww2.amstat.org/meetings/ices/2012/papers/302211.pdf).

[2] Emmanuel Gros, « Études et simulations relatives à la nouvelle méthode de coordination des échantillons d'enquêtes entreprises et établissements développée à l'Insee », Actes des 12e Journées de méthodologie statistique, 31 mars – 02 avril 2015, INSEE.

[3] Anaïs Levieil-Guillon, « Mise en œuvre et résultats de la nouvelle procédure de coordination des échantillons des enquêtes auprès des entreprises et/ou des établissements », Actes des 12e Journées de méthodologie statistique, 31 mars – 02 avril 2015, INSEE.

[4] Arnaud Fizzala et Olivier Sautory, « La coordination négative d'échantillons », Lettre du SSE n° 66, 1er semestre 2011, document interne INSEE.

[5] Circulaire du 16 octobre 2015 relative à la mise en oeuvre des mesures de simplification administrative en matière d'enquêtes statistiques pour les petites entreprises, Bulletin officiel de l'administration centrale n° 2015-11 du 5 novembre 2015, http://www.economie.gouv.fr/bulletin-officiel/boac2015-11.

Département des méthodes statistiques
Version n°1, diffusée le 10 octobre 2017

Insee
Mesurer pour comprendre