Partial Renewal of Business Samples

Arnaud Fizzala

Résumé — Where business surveys are repeated over time and where the focus is on what has changed, the sample is generally not renewed in full each time the survey is conducted, but only in part - half, a third, a quarter, etc. - by way of a trade-off between the response burden deemed to be tolerable for the businesses and the importance attached to the evaluation of changes [1] n the objectives of the survey. This is called a rotating sample in the sense that the part of the sample to be renewed changes ("rotates") each time the survey is carried out. Two techniques are used at INSEE to manage the drawing of these rotating business samples [2] : the Hexal number technique is applied during "normal operation", i.e. when there have been few changes (in scope, sample size, etc.) between two renewals, while the positive coordination technique is used instead in the event of a significant change to the scope of the survey or the sampling plan.

I. UNDER NORMAL CIRCUMSTANCES : HEXAL NUMBER TECHNIQUE

The technique most commonly used at INSEE to manage business sample rotations is based on a fixed number – known as a hexal number – randomly assigned to each sampling frame unit corresponding to the first edition of the survey. The most common case, in which half of the sample is renewed each time, is described below ; however, the method can also be easily applied to renewals of a third, a quarter, a fifth and a sixth of the sample. In order to renew half of a sample, the sampling frame for the first year is divided into two equal parts by randomly assigning each unit a fixed "hexal" number of between 1 and 60 [3]. The units that have been assigned an odd hexal number form the first part of the sampling frame and the units that have been assigned an even hexal number form the second part [4]. he sample for the first year is also effectively split [5] into two sub-samples of equal size.

As a result, for the routine management of the sample in a given year N, the renewal of half of the units in the sample consists of :

> ► on the one hand, retaining one of the two parts of the sampling frame (referred to as the retained part) for year N-1, with units no longer included in the scope
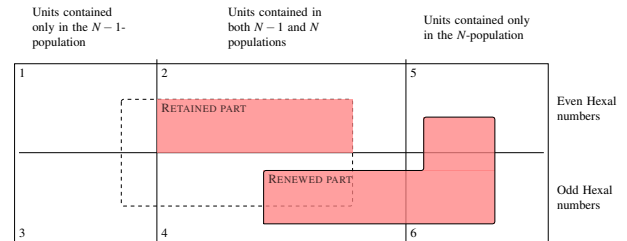

Fig. 1.   Renewal by half using odd hexal numbers impairs

of the survey for year N removed, together with the corresponding part of the sample (referred to as the retained part of the sample) ;

> ► and on the other hand, drawing a new sample from a sampling frame comprising units from the other part of the sampling frame (referred to as the renewed part) for year N-1, also with units no longer included in the scope of the survey for year N removed, to which the new units are added, i.e. included in the scope of the survey for year N. These new units are randomly assigned a hexal number, which may be odd or even and which will be retained for future sampling.

In Figure 1, the sample for year N is made up of all of the pink parts. The sampling actually carried out during year N only involved the units with odd hexal numbers present in the sampling frame for both year N-1 and year N – part 4 –, together with the new units, i.e. those that were not included in the sampling frame for year N-1, regardless of whether their hexal number is even or odd – parts 5 and 6. These units that are actually drawn during year N therefore correspond to the "pink lying-down L shape" in Figure 1 and can be extrapolated to areas 4 + 5 + 6. The units corresponding to the retained part of the sample for year N can be extrapolated [6] to area 2.

In order to reduce [7] the statistical burden on the businesses, this method is generally used by negatively coordinating [2] the part of the sample renewed in year N with [8]the sample from previous years. In cases where the sample is renewed by halves, the aim is therefore to survey the units for two years, but no longer.

---

1. The more the samples overlap, the more accurate the estimate of a change will be.

2. Several household surveys are also based on rotating samples, for example the Rents and Charges survey, for which one fifth of the sample is renewed each quarter and the Statistics on Income and Living Conditions (SILC) survey, for which one ninth of the sample is renewed each year. The methodology used to draw these rotating samples is not the same as the methodology described in this note, which is currently only applied for business surveys.

3. The number 60 was chosen because it is divisible by all the integers between two and six : a hexal number between 1 and 60 therefore allows the sampling frame to be split into two, three, four, five or six parts, thereby enabling the management of renewals by halves, thirds, quarters, fifths or sixths.

4. It is also possible to choose to form the first part using the units that have been assigned hexal numbers between 1 and 30 and the second part from units that have been assigned hexal numbers between 31 and 60.

5. Based on the parity of the hexal numbers of the constituent units.

6. It should be noted that the sampling weight of a unit within the retained part is derived from the sampling performed in year N-1 and must therefore also be retained, even if the characteristics of the unit have changed and would have placed it in a different stratum had sampling taken place in year N.

7. Or, in this case, improved distribution of the statistical burden between businesses that are "statistically equivalent".

8. Within the scope of the general negative coordination of the business samples, the renewed part is also negatively coordinated with the samples from other recent business surveys, but efforts are first made to minimise overlap with the sample used for the previous edition of the survey. Technically, this is achieved by applying a lower "burden parameter" to the other surveys during the performance of the negative coordination.

II. Under Exceptional Circumstances : Positive Coordination

It should be noted that, in the event that substantial changes are made to the sampling plan that have a significant impact on the sample, the hexal number method described above is not suitable for partial renewal of the sample, since it is impossible to change the retained part of the sample. For example, in the extreme case of the sample size being halved in a given year N, the renewed part of the sample should be limited to just a few units (or none if there were no units entering or leaving the scope of the survey), thereby creating an imbalance in the sizes of the two parts of the sample in subsequent years. Where such problematic cases arise, a more "flexible" technique, known as positive coordination, is used to partially renew the sample. This technique uses the same tools as the negative coordination technique, but with the opposite objective, i.e. aiming for maximum overlap between samples.

The biggest disadvantage posed by the positive coordination technique is the difficulty in controlling the overlap between the samples and the length of time the units have been in the panel. It is for this reason that the hexal number method is generally preferred where changes to the survey have not had a significant impact on the sample. Nevertheless, with the implementation of sampling at business level in the coming years, this rule is likely to change. Indeed, as the contours of businesses change from one year to the next, there are several questions [9] that will need to be answered in order to continue using the hexal number method, which was designed to be used with sampling frames that remain stable over time.

III. Examples of Use at INSEE

By way of an example, each year, the INSEE Surveys division draws the sample of establishments for the annual survey on industrial energy consumption (EACEI) and the sample for the business component of the survey on information and communication technologies (Business ICT), both of which have half of their sample renewed each year using the hexal number technique. The Surveys division also draws the sample of microenterprises [10] for the survey on labour activity and employment conditions (ACEMO - TPE), a quarter of the sample for which is renewed each year using the hexal number technique.

The positive coordination method for samples was used, for example, in 2008 when the annual business surveys (EAE), for which half of the sample was renewed each year using the hexal number technique, were replaced by the annual sectoral survey (ESA). The scope of the survey was revised at that time and the size of the non-exhaustive part of the sample was halved. The hexal number method did not appear to be a viable option in view of those changes, so the positive coordination method was used instead [11]. Since the sampling plan had stabilised, with effect from 2009, the renewal of half of the ESA sample using the hexal number technique was reinstated with the sample being managed in this way until 2015. In 2016, positive coordination was once again used in order to manage the transition to the concept of enterprise for sampling.

References

[1] Demoly E., Fizzala A., Gros E. (2014). Méthodes et pratiques des enquêtes entreprises à l'Insee. *Journal de la Société Française de Statistique*, 155, 4, 134 à 159.

[2] Fiche méthodologique "La coordination des échantillons des enquêtes auprès des entreprises".

[3] Cotton F., Hesse C. (1992) Tirages coordonnés d'échantillons. *Document de travail E9206*, Insee.

Département des méthodes statistiques
Version nº 1, diffusée le 24 janvier 2018

---

9. How to manage a change to the contour of a business belonging to the retained part of the sampling frame, for example. Particularly if it absorbs another unit within the retained part.

10. Between 1 and 9 employees.

11. At the time, this was the old method for coordinating samples [3], since the method currently in use was not yet operational.