This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Sample Co-ordination

Contents

General section

1.   Summary

Main objectives of sample co-ordination are to obtain comparable and coherent statistics, high precision in estimates of change over time and to spread the response burden evenly among the businesses. Sample co-ordination means to introduce dependence between two consecutive samples for the same survey or between samples for different surveys in order to minimise or maximise their overlap (number of units in common). Sample co-ordination is very useful and therefore commonly used among business surveys although sample co-ordination makes sampling and estimation more complicated (in contrast to the use of independent samples) because standard methods for sampling and estimation cannot be used in many cases. There are two main categories of methods that can be used for sample co-ordination and within each category there are a number of different methods. Many countries have implemented sample co-ordination but the specific method varies. This module includes an introduction to sample co-ordination in general, main principles to obtain different kind of co-ordination and some comments on sample co-ordination in practice.

2.   General description

The overlap (number of units in common) between samples for different surveys (or between two consecutive samples for the same survey) is random when the samples are drawn independently of each other. In order to have some control over the overlap some kind of method for sample co-ordination can be used. Sample co-ordination can be used to increase the precision in estimates of change over time; the co-ordination design ensures that consecutive samples for the same survey are overlapping, although each sample is drawn from an up-to-date register. Sample co-ordination can also be used to obtain an even distribution of the response burden among the businesses[1]. Businesses often perceive participating in a survey (i.e., completing questionnaires) as a burden because it takes time and effort. The opportunity to spread the response burden among the businesses and improve the precision in estimates of change makes it worthwhile to consider implementing a system for sample co-ordination. Even though sample co-ordination often means that standard methods for sampling and estimation cannot be used.

Several National Statistical Institutes (NSIs) have system and methodology implemented for sample co-ordination but system and methodology varies between the countries. The specific method implemented depends on kind of statistics to be produced, on conditions related to the production of the statistics and on what (or which) objectives of the sample co-ordination the country is focused on. For most of the methods it is easy to show that the given co-ordinated sample is a strict probability sample. However, this can be the case when there is a strong focus on response burden. This is due to the fact that strong focus on response burden often means guaranties on maximum number of surveys a specific business must participate in and/or maximum number of years a specific business must participate in the same survey. There are even sample co-ordination methods that intentionally do not give a strict probability sample in favour for the advantage to have better control over the overlap. Note that sample co-ordination introduces dependence between the obtained samples.

---

[1] The word "business" is used as a generic name for all unit types used in business surveys

Sample co-ordination is commonly used for business surveys but when it comes to individual and household surveys it is more unusual. This is probably related to differences between business and social statistics, differences like the very skew business population and that many countries have a business register (compared to a population register). In addition, the National Accounts has a clear need for comparable and coherent economic statistics in order to compile the Gross Domestic Product (GDP). Sample co-ordination can be obtained in two dimensions: between different surveys and over time for the same survey.

2.1 Sample co-ordination between surveys

Negative sample co-ordination between surveys means that samples for two negatively co-ordinated surveys have as few businesses in common as possible. Entirely successful negative co-ordination means of course completely separate samples. However, there are not always enough businesses to obtain complete negative co-ordination, but this sample co-ordination at least reduces the number of businesses in common. How well negative co-ordination works depends to a large extent on the size of the sample fractions in the different surveys. Note that negative sample co-ordination cannot give guaranties on maximum number of surveys a specific business must participate in when strict probability samples are used. For more information on negative co-ordination in practice, see section 2.7 further down.

Positive sample co-ordination between surveys means that samples for two positively co-ordinated surveys have as many businesses in common as possible. Positive co-ordination between surveys can be used to facilitate comparisons between variable values on the micro level even though data are collected in different surveys. However, this facility is needed most for editing purposes among large businesses and they are almost always completely enumerated and therefore included in all samples independently of co-ordination. This kind of co-ordination can also facilitate the production of comparable and coherent statistics required by (at least) the National Accounts using results from the majority of the economic surveys as building bricks when compiling the GDP. But for small businesses this kind of co-ordination could mean an unnecessary burden taking into account that coherence analysis is most focused on large businesses. And, the requirement from the National Accounts on comparable and coherent statistics can, to a large extent, be met by another kind of co-ordination, namely co-ordination of frame populations; see theme module "Statistical Registers and Frames – Survey Frames for Business Surveys" for more information.

2.2 Sample co-ordination over time for the same survey

Positive co-ordination over time for the same survey is used to obtain high precision in estimates of change over time. This can be achieved when consecutive samples for the same survey are overlapping, i.e., two consecutive samples have many businesses in common. The size of the overlap is stochastic and depends to a large extent on the sampling design, on sampling fractions as well as on changes in the business population between the two sampling occasions. Due to this positive co-ordination over time, a selected business may have to participate in a survey for many years. In order to spread the response burden among the businesses many countries have implemented some kind of method of sample rotation, for more information see section 3.2.

2.3   Co-ordination of surveys based on different kind of units

There is a third kind of co-ordination, not mentioned before, namely co-ordination between business surveys based on different kind of unit types. The fact that business surveys use different kind of statistical units in the BR to construct a frame population and to draw a sample implies a need for this third kind of co-ordination. There are methods for sample co-ordination that admit co-ordination between surveys based on different kind of unit types. The fact that the business population changes very fast in terms of registrations, de-registrations, mergers, split-offs, breakups and take-overs makes it a challenge to achieve a strong co-ordination lasting over time, especially when it comes to co-ordination of surveys based on different kind of units types. For more information on co-ordination of surveys based on different kind of units, see the method module "Sample Selection – Assigning Random Numbers When Co-ordination of Surveys Based on Different Unit Types is Considered".

2.4   Methods for sample co-ordination

There are in principle two categories of sample co-ordination methods:

1)   Sample co-ordination methods based on Permanent Random Numbers (PRNs)

2)   Sample co-ordination methods *not* based on PRNs

The most common method to obtain sample co-ordination is based on the use of PRN. The basic idea is to associate an independent and unique random number, uniformly distributed over the interval (0,1), with every unit in the register. For every unit persisting in the register the same random number is used on each sampling occasion. In this way we always get a new sample from the updated register but a large overlap with the latest sample can be expected. Every new unit (births) is assigned a new random number while closed-down units (deaths) are withdrawn from the register with their random numbers.

Several countries use sample co-ordination based on PRNs and there exists many variations of this method. In the method modules listed below two different methods for sample co-ordination (based on PRN) are presented:

- "Sample Selection – Sample Co-ordination Using Simple Random Sampling with Permanent Random Numbers"

- "Sample Selection – Sample Co-ordination Using Poisson Sampling with Permanent Random Numbers"

Sample co-ordination methods *not* based on PRNs are in general not used in the NSIs. Nevertheless, one of those methods, which is based on linear programming, can be applied in business surveys (Reiss et al., 2003). A main feature with methods based on linear programming is the possibility to optimise (maximise or minimise) the overlap between two samples, two consecutive samples for the same survey or two samples for different surveys. See Ernst (1996, 1998, 1999 and 2002) for a more general description of non-PRN methods.

2.5   Estimation of variances when samples are co-ordinated by using PRNs

In the area of economic statistics measures of change are key parameters and it is of great importance to be able to determine whether an observed change is statistically significant or not. Sample co-ordination by PRNs makes the level estimates correlated and this correlation is quite complicated to

estimate because the use of PRNs brings an additional component of randomness to the rotation pattern (compared to ordinary panel design). The problem of estimating the variance for measures of change when samples are co-ordinated by PRNs has been addressed by several persons during the years. However, a complete and workable method for estimating this correlation under the Statistics Sweden sampling method (co-ordinated SRS and stratified SRS based on PRN) was developed in the late 1990's, see Nordberg (2000). This approach can hopefully be of interest in the context of other PRN systems. For more information on variance estimation, see theme module "Quality Aspects – Quality of Statistics".

2.6   Survey feedback into the Business Register when samples are co-ordinated

Survey feedback means that information obtained from a survey is used to update the Business Register (BR). It is not advisable to use survey feedback from co-ordinated sample surveys, especially not where positive co-ordination over time is used. This applies mainly to variables used in the survey design, variables like economic activity, number of employees and annual turnover. Information on contact variables is not equally sensitive to survey feedback.

Survey feedback implies that businesses in the BR, which are included in samples, are updated, while those not included are not updated. Furthermore, the large overlap between two consecutive samples means that the latest sample is based on a BR-version where businesses included in the previous sample are more updated compared to the rest of the businesses in the BR. Survey feedback will in this case lead to bias in the estimates because the sample is no longer representative for the whole frame population. But from this point of view it is all right to update large businesses with survey feedback because they are, almost always, completely enumerated. Nevertheless, in practice there is a strong desire to be able to use all information collected from surveys to update business information in the BR. During the years work has been done in order to estimate the magnitude of the introduced bias as well as on methods to reduce this bias in the estimation phase. However, at the moment there is no complete and workable method to recommend. In general, feedback from co-ordinated sample surveys should rather be used as quality indicators in the maintenance of the BR.

2.7   Negative sample co-ordination in practice

In recent years there has been a strong focus on lowering the response burden in all EU-countries. There is a challenge in reducing the response burden without negatively affecting the quality of the desired estimates. Sample co-ordination is a successful method only if the number of businesses among which the response burden is spread is sufficient. This is not always the case because the structure of the business population is generally very skewed, consisting of a huge number of small enterprises and a very small number of medium and large-sized enterprises (size in terms of persons employed or other economic variables). Table 1, below, shows the structure of the business population in the EU-countries (source: *Key figures on European business with a special feature on SMEs*). The division into size classes is based on number of persons employed.

Table 1. Total number of enterprises and their distribution by size class among EU countries

	Total number of enterprises (thousands)	Distribution of enterprises by size class			
		Micro < 10 %	Small 10- <50 %	Medium 50- <250 %	Large 250- %
EU-27	*20 994*	*92,0*	*6,7*	*1,1*	*0,2*
Belgium	426	92,5	6,3	0,9	0,2
Bulgaria	270	88,7	9,2	1,9	0,3
Czech Republic	899	95,1	3,9	0,8	0,2
Denmark	211	85,0	12,2	2,4	0,4
Germany	1 880	83,0	14,1	2,4	0,5
Estonia	46	83,9	13,0	2,7	0,4
Ireland	158	87,8	9,9	1,9	0,3
Greece	:	:	:	:	:
Spain	2 653	93,1	6,0	0,8	0,1
France	:	:	:	:	:
Italy	3 947	94,3	5,1	0,5	0,1
Cyprus	47	92,3	6,4	1,1	0,2
Latvia	70	84,4	12,9	2,4	0,3
Lithuania	139	88,7	9,2	1,9	0,3
Luxembourg	17	85,8	11,5	2,2	0,5
Hungary	566	94,3	4,7	0,8	0,2
Malta	:	:	:	:	:
Netherlands	583	90,4	8,0	1,4	0,3
Austria	294	87,2	10,8	1,7	0,4
Poland	1 556	95,5	3,3	1,0	0,2
Portugal	778	94,0	5,1	0,7	0,1
Romania	506	88,9	8,8	1,9	0,4
Slovenia	93	92,4	6,1	1,3	0,3
Slovakia	59	71,2	24,2	3,7	0,9
Finland	202	91,7	6,9	1,1	0,3
Sweden	586	94,7	4,4	0,8	0,2
United Kingdom	1 731	89,3	8,8	1,5	0,4

Table 1 shows that all EU-countries have a skewed business population. Countries with a relatively small total number of businesses often meet a more difficult situation when it comes to spreading the response burden compared to countries with a large total number of businesses. This is due to the fact that quality (in terms of standard errors) in desired estimates is mainly correlated with the sample size (and not with the population size). And, even though total number of businesses is large a detailed stratification can lead to large sampling fractions in specific strata. On the other hand, detailed stratification is often needed in order to produce detailed domain estimates of high quality.

Negative co-ordination is a very effective tool to spread the response burden among small businesses. This is important because they often do not have the capacity to participate in many surveys. However, there is little room for spreading the response burden among medium-sized businesses because these businesses are few, see table 1. In addition, they have a proportionately large impact on the estimates in terms of economic variables and they must therefore often be included in samples. Medium-sized businesses could meet a heavy burden, especially in industries with few businesses. The small number of large businesses, see table 1, are of great importance for the economic statistics because they have a

large impact on the estimates in terms of economic variables. Therefore it is crucial, with few exceptions, to include all large sized businesses belonging to the frame population for a specific survey. Otherwise it would be more or less impossible to publish the survey results. The work on response burden regarding large sized businesses must mainly focus on simplifying for the respondents to supply the requested information.

The possibility of spreading the response burden depends of course, to a large extent, on the structure of the business population in a specific country. The structure is almost always skewed but if the total number of businesses (all categories) is considerably large the possibility for successful negative sample co-ordination increases.

3.   Design issues

3.1   Considerations before introducing sample co-ordination

It is very important to consider some kind of optimal co-ordination (negative as well as positive) between surveys before the sample co-ordination is introduced into a system. Once a survey is placed into the system it is preferable to keep the survey in the same place because of the desired overlap between two consecutive samples for the survey. There are mechanical, as well as manual, methods that can be used to place surveys into a system. However, a number of well-considered decisions must be taken to obtain the best solution in the long run. This could be considerations like:

- optimal co-ordination between the surveys, how are the surveys related to each other

- economic activities and size classes covered by the survey

- periodicity of the survey (monthly, quarterly, annual, periodic)

- time point when questionnaires for different surveys are sent out

- survey content, in terms of number and type of variables

3.2   Sample rotation

As mentioned before, due to the positive co-ordination over time, a selected business may have to participate in a survey for many years. In order to spread the response burden among the businesses it is possible to implement some kind of sample rotation into the sample co-ordination system. The objective of this sample rotation is to keep a selected business in the sample for a pre-specified number of years and then let it rotate out of the sample (in contrast to the stochastic rotation obtained by changes in the business population). There are several methods to obtain sample rotation and Ohlsson (1995) gives a description of some of the methods. As mentioned before, a sample rotation method cannot give guaranties on maximum number of years a specific business has to participate in the same survey when strict probability samples are used.

The number of years a business should participate in a survey is a balance between response burden and the decrease in the precision of the estimates of change over time that is acceptable. And in practice, rotation works only successfully if there is room for rotation. And by successful rotation is meant that a business can rotate out of a sample after the pre-specified number of years without immediately rotating into the sample of another survey. Such room is only available among small businesses where the sampling fraction is small. It takes longer time for businesses in stratum with

larger sampling fraction to rotate out of the sample. How long depends to a large extent on the size of the sampling fraction.

4. Available software tools

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Ernst, L. R. (1996), Maximizing the overlap of sample units for two designs with simultaneous selection. *Journal of Official Statistics* **12**, 33–45.

Ernst, L. R. (1998), Maximizing and minimizing overlap when selecting a large number of units per stratum simultaneously for two designs. *Journal of Official Statistics* **14**, 297–314.

Ernst, L. R. (1999), The maximization and minimization of sample overlap problems: a half century of results. In: *Proceedings of the International Statistical Institute*, 52nd Session, Finland, 168–182.

Ernst, L. R. and Paben, S. P. (2002), Maximizing and minimizing overlap when selecting any number of units per stratum simultaneously for two designs with different stratifications. *Journal of Official Statistics* **18**, 185–202.

Key figures on European business with a special feature on SMEs. Eurostat pocketbooks, ISSN 1830-9720.

Nordberg, L. (2000), On Variance Estimation for Measures of Change When Samples are Coordinated by the Use of Permanent Random Numbers. *Journal of Official Statistics* **16**, 363–378.

Ohlsson, E. (1995), Coordination of samples using permanent random numbers. In: *Business Survey Methods* (eds. Cox, B. G., Binder, D. A., Chinnapa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S.), Wiley, New York, Chapter 9, 153–169.

Reiss, P., Schiopu-Kratina, I., and Mach, L. (2003), The use of the transportation problem in coordinating the selection of samples for business surveys. In: *Proceedings of the Survey Methods Section*, Statistical Society of Canada, Annual Meeting, June 2003.

Interconnections with other modules

8.   Related themes described in other modules

1.   Repeated Surveys – Repeated Surveys

2.   Statistical Registers and Frames – Survey Frames for Business Surveys

3.   Weighting and Estimation – Main Module

4.   Quality Aspects – Quality of Statistics

9.   Methods explicitly referred to in this module

1.   Sample Selection – Sample Co-ordination Using Simple Random Sampling with Permanent Random Numbers

2.   Sample Selection – Sample Co-ordination Using Poisson Sampling with Permanent Random Numbers

3.   Sample Selection – Assigning Random Numbers When Co-ordination of Surveys Based on Different Unit Types is Considered

10.   Mathematical techniques explicitly referred to in this module

1.

11.   GSBPM phases explicitly referred to in this module

1.

12.   Tools explicitly referred to in this module

1.

13.   Process steps explicitly referred to in this module

1.

Administrative section

14. Module code

Sample Selection-T-Sample Co-ordination

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	28-02-2013	first version	Annika Lindblom	Statistics Sweden
0.2	30-04-2013	improvements based on the Norwegian and Swiss reviews	Annika Lindblom	Statistics Sweden
0.3	29-05-2013	improvements based on the Norwegian and Swiss reviews	Annika Lindblom	Statistics Sweden
0.3.1	18-09-2013	preliminary release		
0.4	27-09-2013	improvements based on the EB-review	Annika Lindblom	Statistics Sweden
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:43