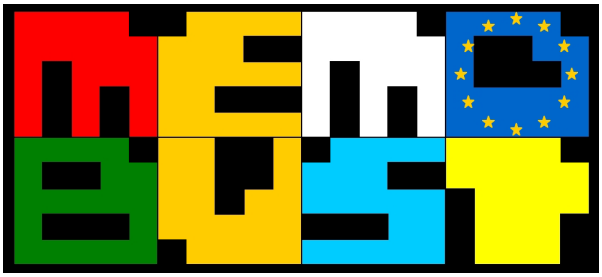


This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Method: Assigning Random Numbers When Co-ordination of Surveys Based on Different Unit Types is Considered

Contents

General section

1. Summary

Sample co-ordination by the use of Permanent Random Numbers (PRNs) is a common method used to have some control over the overlap (number of businesses in common) between samples for two different surveys or between consecutive samples for the same survey. The basic idea is to associate an independent and unique random number, uniformly distributed over the interval (0,1), with every unit in the Business Register. A BR generally consists of several unit types and unit type for a business survey is chosen on the basis of the statistics to be produced. This means that all unit types must be assigned PRNs. There are various methods for this but the most straightforward way would be to assign PRNs to each unit type separately. This method means that samples based on different unit types are independent but it does not admit co-ordination between such surveys. The fact that business surveys use different unit types implies a need for this kind of co-ordination. Especially the possibility of negative co-ordination between surveys based on different unit types (in order to spread the response burden) is important when it comes to small businesses.

Another approach to assign PRNs would be to use a method implying that the unit types can be co-ordinated through the PRNs. This method has the advantage to admit sample co-ordination between unit types but, as a drawback, brings dependence between samples based on different unit types. Co-ordination through PRNs cannot meet all objectives of sample co-ordination equally strong and different strategies are discussed in more detail below and references are given to other parts of the handbook.

2. General description of the method

2.1 Co-ordination when several unit types in the Business Register are considered

Sample co-ordination can be used to have some control over the overlap (number of businesses in common) between samples for two different surveys or between consecutive samples for the same survey. The 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 among the businesses[1], see theme module "Sample Selection – Sample Co-ordination" for more information. A common method to obtain sample co-ordination is based on the use of Permanent Random Numbers (PRNs). The basic idea is to associate an independent and unique random number, uniformly distributed over the interval (0,1), with every unit in the Business Register (BR). The method modules "Sample Selection – Sample Co-ordination Using Simple Random Sampling with Permanent Random Numbers" and "Sample Selection – Sample Co-ordination Using Poisson Sampling with Permanent Random Numbers" give different examples of sample co-ordination based on PRNs. The present module discusses assigning PRNs when several unit types in the BR are considered.

The majority of the National Statistical Institutes (NSIs) have not implemented co-ordination of surveys based on different unit types but Australia, France and Sweden are examples of countries using this kind of co-ordination. Australian Bureau of Statistics (ABS) achieves co-ordination between

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

samples of different types of units by the way the PRNs are assigned; see Brewer et al. (2000) for more information. The method used at ABS is quite similar to the method used in Sweden.

Institut National de la Statistique et des Études Économiques (INSEE) uses a somewhat different method to co-ordinate samples of different types of units. The co-ordination between unit types is mainly obtained by the way lower level units are connected to their higher level linked unit. See Hesse (1999) for more information.

The methodology described in this module is used in Statistics Sweden's system for co-ordination of frame populations and samples from the Business register (SAMU). For a general description of SAMU see Lindblom (2003).

A BR generally consists of several unit types and each business survey chooses unit type in accordance with the statistics to be produced. For example, institutional statistics is generally based on the enterprise unit, functional statistics is generally based on the kind of activity unit and regional statistics is generally based on the local kind of activity unit. Two types of sample co-ordination are commonly used (discussed in theme module "Sample Selection – Sample Co-ordination"), namely 1) co-ordination over time for one specific survey and 2) co-ordination between surveys based on the *same* unit type. However, there is a third kind of co-ordination to consider, namely co-ordination between surveys based on *different* unit types.

2.2 Assigning random numbers when several unit types in the BR are considered

The fact that business surveys use different kind of units in the BR means that all unit types must be assigned PRNs. There are several methods but the most straight-forward method would be to assign PRNs to each unit type separately meaning that the set of PRNs assigned to one unit type is completely independent of the set of PRNs assigned to another unit type. This method is simple and has the advantage that samples based on different unit types are independent of each other. However, it does not admit sample co-ordination between surveys based on different unit types. This drawback affects especially small businesses where the possibility to co-ordinate negatively (to spread the response burden) between surveys based on different unit types is very important.

Another approach to assign PRNs would be to use a method implying that the unit types can be co-ordinated through the PRNs. This method has the advantage to admit sample co-ordination between unit types but, as a drawback, brings dependence between samples based on different unit types. In the simple case with single-location and single-activity businesses this method means to assign the same random number to all units within a business. And, the majority of the small businesses consist of single-location and single-activity businesses which means that the proposed method for co-ordination between unit types works very well in this case. For the multiple-location and/or multiple-activity businesses this kind of co-ordination is less efficient because it is only possible to co-ordinate a multiple-location and/or multiple-activity enterprise with *one* of its lower level linked units. However, the majority of these businesses are large and large businesses are almost always included in samples so there are limited opportunities for spreading the response burden among them.

2.3 Principles for co-ordination

Co-ordination through PRNs offers a simple way to obtain co-ordination between unit types even though this method cannot meet all three objectives of co-ordination equally strongly. The reason is that the strategy to obtain the different objectives of co-ordination is somewhat contradicting:

- co-ordination over time for one specific survey and co-ordination between surveys based on the same unit type requires PRNs as permanent as possible
- co-ordination between surveys based on different kind of unit types would require PRNs that are, to some extent, updated

Strongest co-ordination, for one specific survey over time and between surveys based on the same unit type, is obtained by keeping the initially assigned PRN as permanent as possible. On the contrary, to maintain a strong co-ordination over time between unit types means that the PRNs needs to be somehow updated in order to follow changes in the business population in terms of registrations, de-registrations, mergers, split-offs, breakups and take-overs. An initially perfect co-ordination between unit types will otherwise gradually degenerate.

Updating PRNs contradicts the requirement from the two other types of co-ordination, namely keeping the PRNs as permanent as possible. To conclude, main objectives of co-ordination must be considered prior to the introduction of a system for co-ordination of surveys by the use of PRNs. Focus only on co-ordination over time for one specific survey and co-ordination between surveys based on the same unit type means that the best method is to assign PRNs to each unit type separately. Focus also on co-ordination between surveys based on different unit types means additional demands on the method for assigning PRNs.

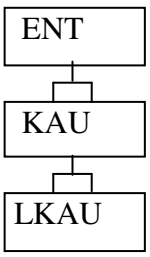
2.4 Unit types in a Business Register

A BR includes several unit types, generally at least the following:

- Enterprise Unit (ENT)
- Kind of Activity Unit (KAU)
- Local Kind of Activity Unit (LKAU)

A BR often includes more unit types compared to the above mentioned but principles for co-ordination of surveys based on different kind of units can easily be applied to a BR-structure including more unit types.

The relationship between the above mentioned unit types are showed in the figure below:



The unit types in the BR are linked together in a hierarchical way. In this example the LKAU is the smallest building brick in the BR. Each LKAU is linked to *one* upper level KAU and several LKAUs can be linked to the same upper level KAU. In the same way, each KAU is linked to *one* upper level ENT and several KAUs can be linked to the same upper level ENT.

2.5 Top-down or bottom-up approach when assigning random numbers

PRNs are assigned to all new units in the BR and a vital question is whether the assignment should be done by a "top-down" or a "bottom-up" approach. A top-down approach would mean to start the assignment of PRNs on the enterprise level and then go further down to the lower level linked units within the enterprise. And consequently, a "bottom-up" approach would mean to start the assignment of PRNs on the LKAU level and then go further up to the higher level linked units within the enterprise. The top-down approach means that a new enterprise is assigned a new random number and that the lower level linked KAU is assigned the same random number. If an enterprise has several lower level linked KAUs, one of them is assigned the same random number as the enterprise. Remaining new KAUs are assigned new random numbers. LKAUs are assigned random numbers according to the same method. A disadvantage with the top-down approach arises when a *new* enterprise is founded by one or more existing lower level linked units. As mentioned earlier, in order to co-ordinate between unit types one lower level linked unit should have the same random as the enterprise. However, one (or more) of the lower level linked units already have a random number and therefore run the risks of being forced to change from the existing random number to the new random number assigned to the enterprise.

The bottom-up approach means that a new LKAU is assigned a new random number and that a *new* higher level linked KAU is assigned the same random number. If a new KAU has several lower level linked LKAUs, the new KAU is assigned one of the LKAUs random number. And accordingly, a new enterprise is assigned the random number from one of its lower lever linked LKAUs. Note that another method (within the bottom-up approach) would be to assign a new enterprise the random number from one of its lower level linked KAUs. Examples 4.1 and 4.2 illustrate the difference between those two methods (or strategies).

The situation where a new enterprise is founded by existing lower level linked units causes no problem when using the bottom-up approach. Although, a disadvantage is that it can cause random number duplicates on the enterprise (and KAU) level due to changes in the business population in terms of mergers, split-offs, breakups and take-overs. However, the problem with random number duplicates can be solved quite easily.

2.6 Assigning PRNs to single-location and single-activity enterprise

In the simple case (single-location and single-activity enterprises) co-ordination of unit types through PRNs means to assign the same random number to all units within the enterprise. Note that this simple case applies to the absolute majority of the enterprises in the BR. In other words; the kind of activity unit and the enterprise unit are assigned the same random number as the local kind of activity unit. Bear in mind that a single-location and single-activity enterprise can change into another more complex structure and to maintain the co-ordination requires well considered continuity rules for the PRNs.

2.7 Assigning PRNs to multiple-location and/or multiple-activity enterprises

The assignment of random numbers to a multiple-location, or multiple-activity, enterprise is more complicated when co-ordination between unit types is considered. There are several possibilities to assign PRNs in this case and, compared to a single location and/or activity enterprise, the co-ordination for a multiple-location and/or multiple-activity enterprise will of course be less efficient. This is due to the fact that it is only possible to co-ordinate a multiple-location and/or multiple-activity enterprise with *one* of its lower level linked units. But the objectives to obtain comparable and coherent statistics imply that the method should facilitate co-ordination of the most important "enterprise-like" units within an enterprise unit. In several countries serves functional statistics as the most important input to the National Accounts. In addition, many other users of economic statistics want to follow different economic activities over time. A way of meeting this requirement would be to give the largest unit (from each unit type) classified into the same industry as the enterprise the same random number. Number of employees/persons employed is, in general, auxiliary information known at each unit type and therefore recommended to use as the size measure. Another approach would be to give the largest unit (from each unit type) classified into the same region as the enterprise the same random number. The chosen method for co-ordination of units within a multiple location/multiple activity enterprise must be decided after taking different demands on co-ordination into account.

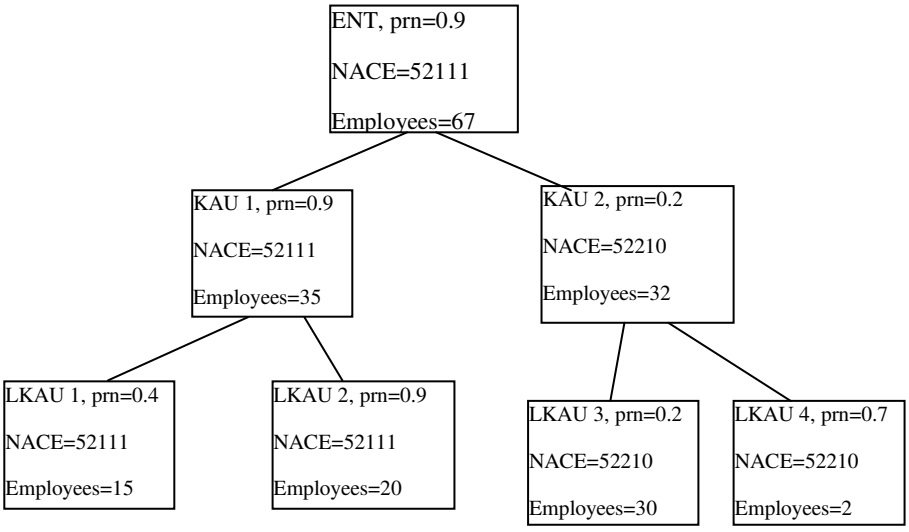
3. Preparatory phase

4. Examples – not tool specific

Different strategies can be used when assigning PRNs according to the "bottom-up" approach:

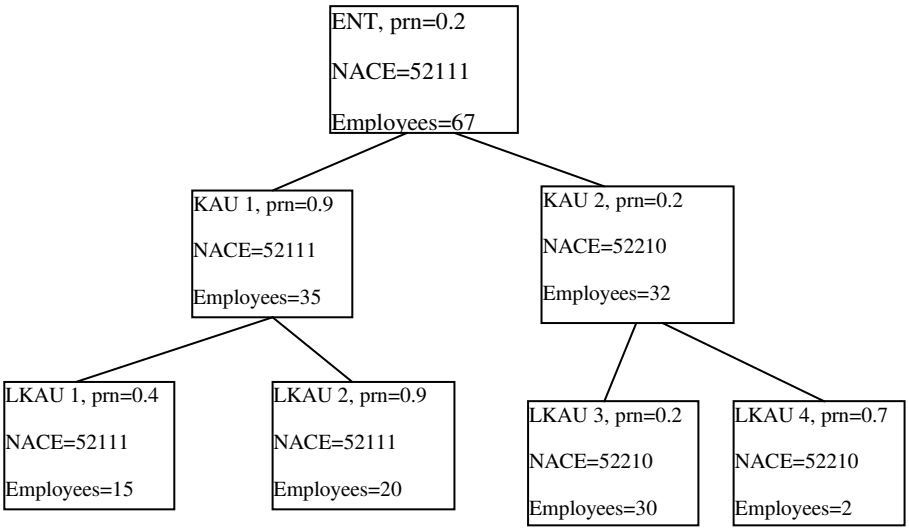
- Strategy *A* means to select a main unit on each level in respect to the closest upper level linked unit, see example 1 below.
- Strategy *B* means to select a main unit on each level in respect to the enterprise unit, see example 2 below.

4.1 Example 1



In the first example PRNs are assigned to each LKAU. Applying strategy A in this example means that, according to the earlier mentioned rules, KAU 1 is assigned the same PRN as LKAU 2 because this LKAU is the largest LKAU classified into the same two-digit industry as the KAU 1. In the same way, KAU 2 is assigned the same PRN as LKAU 3. Keeping to strategy A when assigning a PRN to the enterprise means selecting the main KAU and assign this PRN to the enterprise. This is KAU 1 because this is the largest KAU within the same industry (two digit-level) as the enterprise.

4.2 Example 2



Applying strategy B instead of strategy A gives example 2. As in example 1, KAU 1 is assigned the same PRN as LKAU 2 and KAU 2 the same PRN as LKAU 3. But, when assigning a PRN to the enterprise, strategy B means to select the main LKAU in the same industry (two digit-level) as the enterprise. LKAU 3 is the main unit and following strategy B means to directly assign this PRN to the enterprise (and not go via KAUs).

To conclude, strategies A and B give a different PRN to the enterprise level. In the first example LKAU 2 and KAU 1 are co-ordinated with the enterprise. In the second example LKAU 3 and KAU 2 are co-ordinated with the enterprise.

5. Examples – tool specific

6. Glossary

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

7. References

Brewer, K. R. W., Gross, W. F., and Lee, G. F. (2000), PRN Sampling: The Australian Experience. *ISI Proceedings: Invited Papers, IASS Topics, Helsinki August 10-18, 1999*, 155–163.

Hesse, C. (1999), Sampling co-ordination: A review by country. Technical Report E9908, Direction des Statistique d’Entreprises, INSEE, Paris.

Lindblom, A. (2003), SAMU - The system for coordination of frame populations and samples from the Business Register at Statistics Sweden. Background Facts on Economic Statistics 2003:3, Statistics Sweden.

Specific section

8. Purpose of the method

Co-ordination of surveys based on different unit types

9. Recommended use of the method

1. Co-ordination through Permanent Random Numbers (PRNs) offers a simple way to obtain co-ordination of surveys based on different unit types.
2. Negative co-ordination is a very effective tool to spread the response burden among small businesses. Using this method means that negative co-ordination between surveys based on different unit types works very well for small single location businesses.

10. Possible disadvantages of the method

1. The method used to assign PRNs is more complicated.
2. PRNs on different unit types become dependent by using this method (and samples drawn based on different unit types).

11. Variants of the method

1. Top-Down approach when assigning PRNs in multiple-location and/or multiple-activity businesses.
2. Bottom-Up approach when assigning PRNs in multiple-location and/or multiple-activity businesses.

12. Input data

1. A Business Register

13. Logical preconditions

1. Missing values
    1.
2. Erroneous values
    1.
3. Other quality related preconditions
    1.
4. Other types of preconditions
    1.

14. Tuning parameters

1.

15. Recommended use of the individual variants of the method

1.

16. Output data

1.

17. Properties of the output data

1.

18. Unit of input data suitable for the method

19. User interaction - not tool specific

1.

20. Logging indicators

1.

21. Quality indicators of the output data

1.

22. Actual use of the method

1. This method is implemented in Statistics Sweden's system for co-ordination of frame populations and samples from the Business register (SAMU).

Interconnections with other modules

23. Themes that refer explicitly to this module

1. Sample Selection – Sample Co-ordination

24. Related methods described in other modules

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

25. Mathematical techniques used by the method described in this module

1.

26. GSBPM phases where the method described in this module is used

1. Design phase

2. Data collection phase for frame creation and sampling

27. Tools that implement the method described in this module

1.

28. Process step performed by the method

Administrative section

29. Module code

Sample Selection-M-PRN with Different Unit Types

30. Version history

Version	Date	Description of changes	Author	Institute
0.1	01-04-2013	first version	Annika Lindblom	Statistics Sweden
0.2	16-05-2013	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
0.4	15-08-2013	improvements due to new information on useful references	Annika Lindblom	Statistics Sweden
0.4.1	18-09-2013	preliminary release		
0.5	27-09-2013	improvements due to the EB-reviews	Annika Lindblom	Statistics Sweden
1.0	26-03-2014	final version within the Memobust project		

31. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:44