Business sample co-ordination

Jeļena Voronova[1]

[1]Central Statistical Bureau of Latvia, e-mail: jelena.voronova@csb.gov.lv

Abstract

Most of the samples from a Business Register in Statistics of Latvia are independent from each other and basically not coordinated before. Response burden was the main purpose for starting business samples coordinating in 2011/2012. In fact, there are more than one positive aspect for statistics in final. Samples coordinating is based on Sweden method called SAMU and technique JALES. Method is based on permanent random number that is permanently associated with each Business Register unit. 12 samples with different types and purposes are placed on a one scheme for maximum negative co-ordination between them. No positive co-ordination was required from the head of surveys. In addition, the scheme was made with thoughts to increase and expand it, wherever it would be needed. So there are "places" for more surveys with negative and positive co-ordination opportunity with existing dozen. First real unit rotation of coordinated samples will be implemented in 2012/2013.

Keywords: Coordination, business statistics

1 Introduction

Response burden is actual problem in Business statistics in Latvia since time is money. Several methods are used, as decrease number of questionnaires and length, wide use of administrative data, sample size reducing, survey planning and so on. Main purpose is to find optimal and right way between statisticians eternal problem as estimation accuracy, respondent burden, survey planning, cost minimization...

This paper is concentrated on method of samples co-ordination in Latvia's Business Statistics with aim to decrease respondent burden, without any important changing in surveys implementation. Business statistics samples in Latvia based on data from Business Register (BR) and implement as stratified simple random sample (SSRS). All surveys are independent (if there are no additional terms in query) from each other with personal frame, methods. Mathematical support division (MSD) receives samples queries for long-term (current and next years) and short-term (next year) statistics every year in October. As well as actual list of active enterprise from BR. All samples are ready to be used in early December and they will "work" all next year.

2 Sample co-ordination

2.1 Business register and samples

Sampling of all probability samples is done by MSD staff. Frame population used for calculating samples is based on the BR at Statistics Latvia. In November 2010 the actual sample frame contained 88313 enterprises.

There are several special rules and methods in the BR that let distinguish active unit from non-active. 22 samples queries for long- and short-term statistics were received at the end of 2010 (base data for analyzing). 18 samples were calculated in result (there are surveys with single aim and form, exceptions is length of form). Sub-samples and samples with survey period longer than a year were removed from a co-ordination list, and kept in mind. Only 12 samples could be coordinated in time. All enterprises were broken down into five groups by number of employees. In general, there is census in groups of "large" units (more than 50 employees), no sample and administrative data uses with "small" units (9 employees and less) and "middle" enterprises with employee from 10 to 49 employees surveys by random sample (Table 1). It means that only 10.36% of whole population could be coordinated into 12 samples during one time period.

Table 1: Population breakdown by groups of number of employees

Number of employee	Group	Number of enterprises	Proportion of enterprises, (%)
0-9	0	76677	86,82%
10-19	**1**	**5695**	**6,45%**
20-49	**2**	**3453**	**3,91%**
50-249	3	2090	2,37%
250+	4	398	0,45%
	Total	88313	100,00%

Based on 12 non-coordinated samples, which were calculated in November 2010, we can conclude that 8 is the maximum number of forms that an enterprise received to fill during 2011. The main aims of co-ordination is to decrease maximum number of forms what receive one enterprise and the number of enterprises, which do not receive any form from Central Statistics bureau of Latvia; and also to increase the number of enterprises which receive only 1 to 3 forms.

Table 2: Population distribution by number of forms

| Number of survey forms	Number of enterprise	Proportion of enterprises, (%)
0	50139	56,8
1	10006	11,3
2	4712	5,3
3	2403	2,7
4	1254	1,4
5	1025	1,2
6	644	0,7
7	54	0,1
8	1	0,0
Total survey:	70238	79,5
None:	18075	20,5
Total:	88313	100

2.2 Co-ordination methods

The experience of other countries in sample coordinating was studied, to detect methods and ways that are closer for existing sampling methods in Latvia.

Swedish method (SAMU/JALES)

The technique used by the SAMU to select samples was developed at Statistics Sweden in the late 60's. SAMU generates samples from updated BR. Samples are drawn by so called JALES technique is based on permanent, independent and unique random number (PRN),which is uniformly distributed over the interval (0,1) and permanently associated with every BR unit. PRN birth and death occurs at the same time as its unit birth and death. Each sample is calculated using only PRN. To select simple random sampling, we can take n units from any point in the interval (0,1) and any direction from this point. If there is not enough unit at the end (start) of interval, we can continue selection from the start (end) of it. To co-ordinate two samples with sample size n_1 and n_2 it's only necessary to detect starting point a_1 and a_2 in the interval (0,1) and direction (left or right) to start selecting unit (Figure 1).

Figure 1: Two samples co-ordination
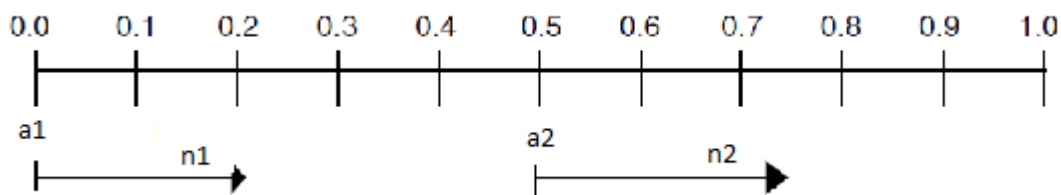
We have to choose the same starting point and direction for both samples to get the maximum positive co-ordination. Select different well apart starting point and use the same directions for negative co-ordination of two samples. If there are not enough units to provide complete negative co-ordination, technique could reduce overlapping.

Unit rotation is needed in SAMU due to the positive co-ordination in time. The last digits in PRN are associated unit to random group. There are five rotation groups in SAMU. One group is rotating in one year by shifting random number (or starting point) on 0.01

Netherland method (EDS)

There are used PRN to co-ordinate samples in Netherland. Method based on burden measurement by estimated time required for complete the form, dividing frame into six classes (Table 3).

Table 3: Class of response burden in EDS

Class	Completion time (min)	Response burden
1	1-30	1
2	31-60	2
3	61-120	4
4	121-180	6
5	181-240	8
6	241+	10

After every sample calculation completion time is accumulated for every unit in the frame and set response class. Before every next sample selection, units are sorted by the PRN and response class (ascending). Thus the most burdened units are sent at the end of the list (Figure 2).

Figure 2: Coordinating samples in EDS after two selections

The file with unit's identification numbers and dummy variables that indicates unit being in the sample for last period was generated for unit rotation. Rotation fraction is calculated taking in account total number of enterprise and possibility to remove unit from next year survey sample. Rotation fraction is in the interval [0;1], where 0 means that next year sample will be maximum positive coordinated with last year sample.

Other methods

France coordination method (OCEAN) was also investigated. It is based on assigning random number, which are recalculated after each sample. Sample selection and unit random methods are similar to Swedish SAMU. Finland method OTKO is relatively young. There is used PRN in this method and response index, equal to 100 at the start. Response index decrease after each unit selection in the sample. When index reduces till determinates level, unit comes out of frame on a time. Then it came back with started index 100.

SAMU method with JALES technique could be used in Statistics Latvia, without global changes in sample design, as well as easy understandable administration and suitable to main guidelines of business survey sample.

2.2 Co-ordination method in statistics Latvia

Optimally best coordinated scheme need to be constructed for Statistics Latvia. All 12 samples should be divided on blocks. Five blocks (Figure 3) is used in SAMU. Left sample direction and positive co-ordination within one year would not use in statistics Latvia.

Figure 3: Five block in the SAMU by Background facts on Economic statistics *et al.* (2003, p.15)
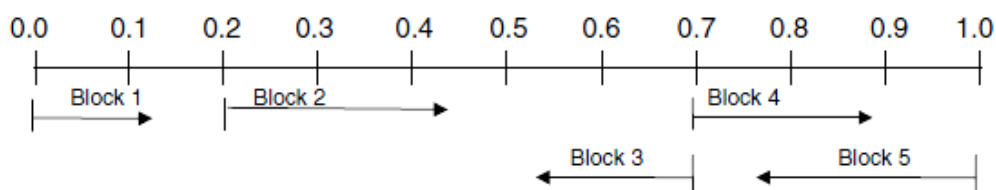
Its own starting point *a* was detected first for every survey in the interval (0,1). After getting results, surveys were grouped and changing its location in several occasions. In practice from twelve samples needed in co-ordination, 7 surveys studies different economics activities (NACE rev2.) and mostly have no overlapping. And there are 5 surveys which target population is mostly whole population of active enterprise (Table 4).

Table 4: Distribution Survey interest by 5 surveys

Number of survey	Number of enterprises	Proportion of enterprises, (%)
0	1	0.0
1	351	3.8
2	431	4.7
3	165	1.8
4	1139	12.5
5	7061	77.2
Total	9148	100.0

Keep in mind only enterprises, with 10 to 49 employees could be coordinated. An optimal scheme (Table 5) was created after swapping block and analyzing result. Best location for every survey was chosen, taking in account:

- Number of questionnaires, that enterprise will receive in time;

- Estimated time required to fill questionnaire.

Each survey got its own place and starting point at co-ordination scheme.

Table 5: Co-ordination scheme for Business Statistics in Latvia since 2011/2012

Block number	1	2	3	4	5	6	7
Surveys	7 studies with no overlapping population frafe	1 survey	1 survey	Free place for rare and special surveys	1 survey	1 survey	1 survey
Starting point	0	0.143	0.286	0.429	0.571	0.714	0.857

Permanent random number is calculated and fixed with rounding 9 digits after decimal point. Unit rotation, after BR advice, will occur in 3 year period. Frame is divided into 3 groups, by the last three digits in PRN. First rotation into frame will occur in November 2012.

Main achievements we have:

- Negative co-ordination between surveys (it is impossible to get entirely non-overlapping samples in situation with small population);

- Positive co-ordination in time;

- Response burden.

References

Background facts on Economic Statistic (2003). SAMU. The system for co-ordination of frame populations and samples from the Business Register at Statistics Sweden. *Statistics Sweden, 2003:3*

Coordination of samples: the microstrata methodology. SALOMON. Eurostat project.

Mészáros, P. A program for sample co-ordination SALOMON, Contributed paper

Nedyalkova, D., Pea, J., Till lé, Y.(2009). A Review of Some Current Methods of Coordination of Stratifed Samples. Introduction and Comparison of New Methods Based on Microstrata

Ohlson, E. (1992) SAMU. The system for Co-ordination of samples from the Business Register at Statistics Sweden. *Statistics Sweden , R&D report 1992:18*

Teikari, I. (2001). Poisson mixture sampling in controlling the distribution of response burden in longitudinal and cross section business survey. *Research reporst* 232, Statistics Finland.

Valliant, R. (2007) Survey sample coordination. A Summer School co-arranged by Orebro University and Statistics Sweden. *University of Michigan, U.S., Joint program in Survey Methodology University of Maryland.*