



Coordinated Sampling for Business Surveys –

theory, method and application at Statistics Sweden

Annika Lindblom
Statistics Sweden



Introduction

Statistics Sweden conducts about 60 business surveys during a year and half of them are part of the economic-statistical system with the National Accounts (NA) as a main user.

The majority of the surveys within the economical-statistical system is part of the so called SAMU-system which means coordination of frame populations and samples.

Business surveys conducted at Statistics Sweden during the year 2017

SAMU	Number of Business Surveys	Number of selected units (LU)	Number of collections
Yes	29	154 950	589 039
No	34	146 450	482 367
Total	63	301 400	1 071 406



Coordination

There are three main objectives of the SAMU-system:

- To obtain comparable statistics – *National Accounts and other users*
- To obtain high precision in estimates of change over time – *coordination over time*
- To spread the response burden / facilitate comparisons of variable values from different surveys – *coordination between surveys (positive or negative)*



Method for sample coordination

The sample coordination method in SAMU is based on:

- Sequential simple random sampling without replacement (sequential srswr)
- Permanent Random Numbers (JALES-technique)



Sequential simple random sampling without replacement

- Each unit in the register is assigned a random number over the interval $(0,1)$
- Order
- *Fix sample size*: let the n first units constitute the sample
- *Random sample size*: let the n first units with random number less than a specific value constitute the sample

It is intuitively clear that the result is a srswor. A formal proof is given in Ohlsson, E. (1992). SAMU – The System for Co-Ordination of Samples from the Business Register at Statistics Sweden. Also included as the last slide in this presentation.

JALES-technique

The basic idea is to:

- associate an independent and unique permanent random number (PRN) uniformly distributed over the interval $[0, 1]$, with every unit in the register
- For every unit in the register persisting in the register the same random number is used on each sampling occasion
- Every new unit is assigned a new random number while closed-down units are withdrawn from the register with their random number

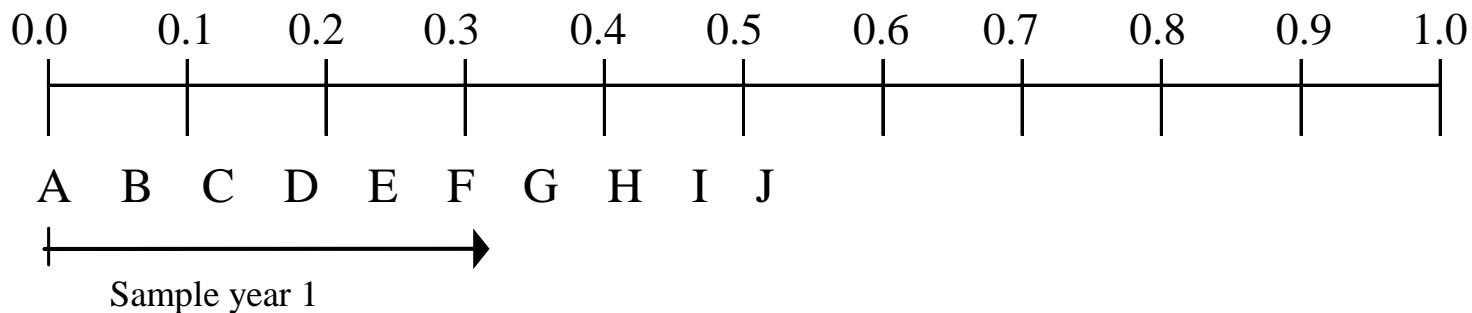
Note that the PRN:s is assigned and maintained in the frozen versions of the BR, not in the up-to-date BR.



Figur 1 Co-ordination over time for the same survey

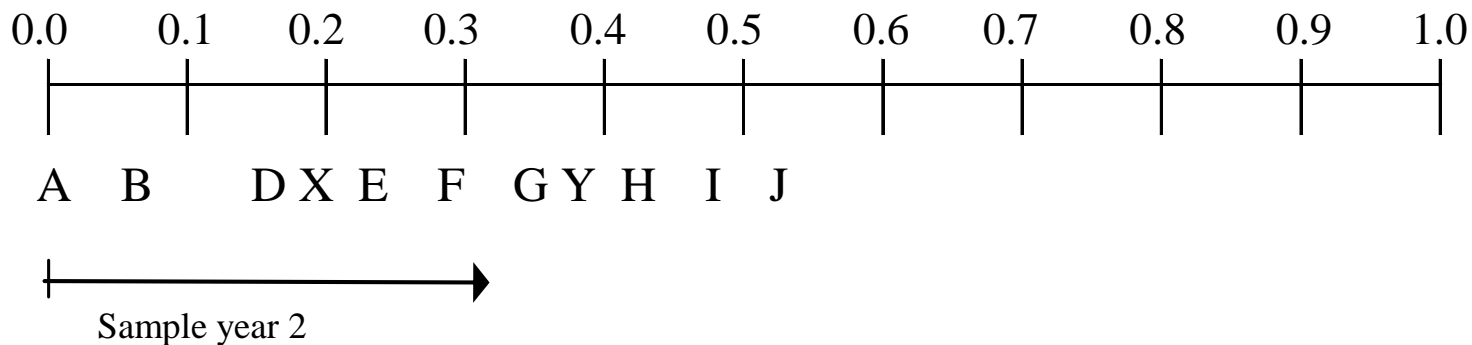
Frame population year 1

Statistics Sweden



Frame population year 2

Statistiska centralbyrån



Sample coordination over time



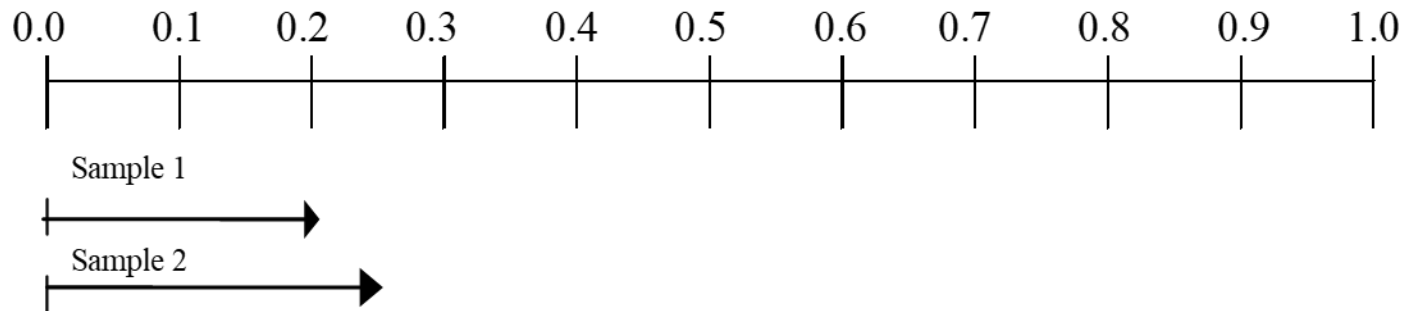
- Two consecutive samples for the same survey are overlapping (although each sample is drawn from an up-to-date BR):
 - Obtain high precision in estimates of change over time
 - Establish good contact with the respondents (reducing measurements errors)
- The size of the overlap is stochastic and depends to a large extent on sampling design, sampling fractions as well as on changes in the business population between the two sampling occasions



Sample coordination between surveys

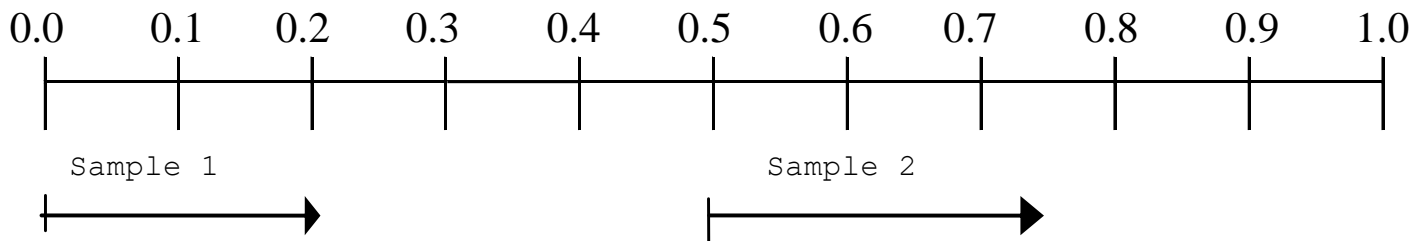
- The uniform distribution is symmetric
- Select the small n first units to the left, or to the right, of any fixed point on the interval $(0,1)$

Figur 2 Positiv co-ordination between two surveys

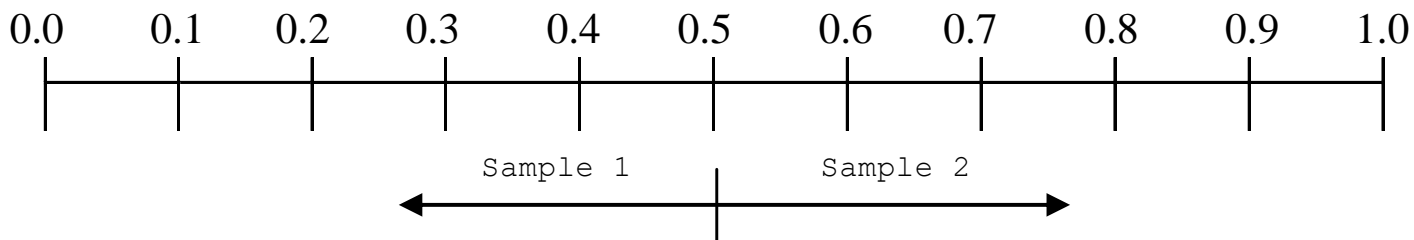


Figur 3 Negativ Co-ordination between two surveys

Statistics Sweden



Statistiska centralbyrån



Sample coordination between surveys

- *Negative coordination*: spread the response burden (samples for different surveys have as few units in common as possible)
- *Positive coordination*: facilitate comparisons of variable values in various surveys (samples for different surveys have as many units in common as possible)

Assigning PRNs

- A BR generally consists of several unit types and all unit types used for statistical purposes must be assigned PRNs
- Most straightforward would be to assign PRNs to each unit type separately
- This means that samples based on different unit types is independent but does not admit coordination between such surveys

Assigning PRNs

- Random numbers are assigned to the smallest building brick in the BR; the local kind of activity unit
- The unit types are then co-ordinated through their random numbers:
 - all units belonging to a single-location (activity) enterprise has the same random number
 - units in a multiple-location (activity) enterprise is given the random number of one of its lower level linked units (the main one in some aspect)
- This means that samples based on different unit types is dependent but does admit coordination between such surveys

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 coordination over time.

Sample rotation

Necessary due to the sample co-ordination over time – otherwise a selected unit may have to participate in a survey for many years

Response burden versus decrease in the precision of estimates of change over time

Objective to keep a selected unit in the sample for five years

Partly renewing the sample

Important that positive and negative coordination for the same survey over time and between surveys is maintained after sample rotation is implemented

Sample rotation

- In Sweden, there is room for rotation in strata with sampling fraction less than 10 percent
- Two common methods for sample rotation:
 - shift the starting point
 - shift the random numbers

The advantage with shifting random numbers is mainly a better overview of how the different surveys is placed into the system, the picture of starting points do not have to be updated every year

Sample rotation

- To rotate a unit in a stratum with sampling fraction less than 0.1 out of the sample after five years means to shift the starting point (or random number) by 0.02 a year
- The disadvantage of shifting the random numbers (or starting points) each year with the same length is that the rate of the rotation will vary considerably among strata.
- Random rotation groups is used to obtain the same rotation grade in all strata (preferable to shift random numbers together with random rotation groups compared to shift the starting points)

SAMU in practice



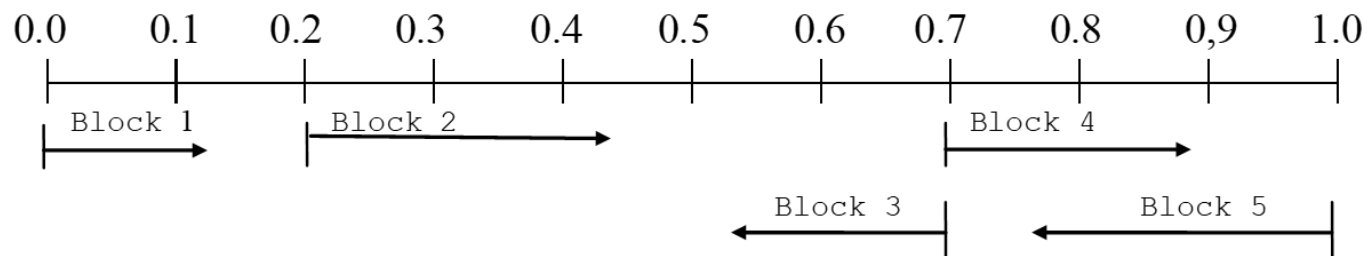
- The SAMU-system is based on four frozen versions of the (BR) each year. Originally based on the frozen version but additional information is provided as well as some re-classification of units
- Those occasions are selected to fit the needs from the surveys as well as the quality of the BR
- Two different sampling designs are currently used in SAMU:
 - (Stratified) simple random sampling without replacement
 - (Stratified) Pareto π ps sampling without replacement

SAMU in practice



- Recommendation:
 - annual surveys referring to the year (t) uses the frozen version of the BR established in November year (t)
 - sub annual (monthly and quarterly) surveys referring to the year (t) uses the frozen version of the BR established in March year (t)
- Improved coverage due to many changes in the business population close to the turn of the year
- Improved possibility for comparisons between annual and sub annual surveys referring to the same year (t)

Sample co-ordination in the SAMU



Note that this figure should not be seen as an ideal placing of blocks but rather as the result of 30 years of additions and adjustments.

SAMU in practice

Administration of SAMU

- Each survey in SAMU exists in a table with associated block (starting point and sampling direction).
- Placing new surveys are done manually, SAMU do not have any methodological assistance when placing a new survey into a block. Common sense is used!

Software for sample selection

- Statistics Sweden has developed a general and flexible tool for sample selection. A tool possible to use for all sorts of frame populations and samples: coordinated, independent, multistage and multiphase drawn from any kind of register.
- The sample selection tool is developed in the software SAS and consist of several SAS-macros, more or less one macro per sampling design.
- However, there are one specific “package” for sample surveys included in the economic-statistical system

Software for sample selection

In current version of the sampling selection tool the following sampling designs are covered (all of them use a sequential sampling scheme and can be used with stratification):

- Simple random sampling
- Bernoulli sampling
- Poisson sampling
- Systematic sampling
- Systematic π ps sampling
- Pareto π ps sampling
- Sequential poisson sampling
- Poisson mixture sampling
- Pareto mixture sampling
- Cluster sampling in two stages

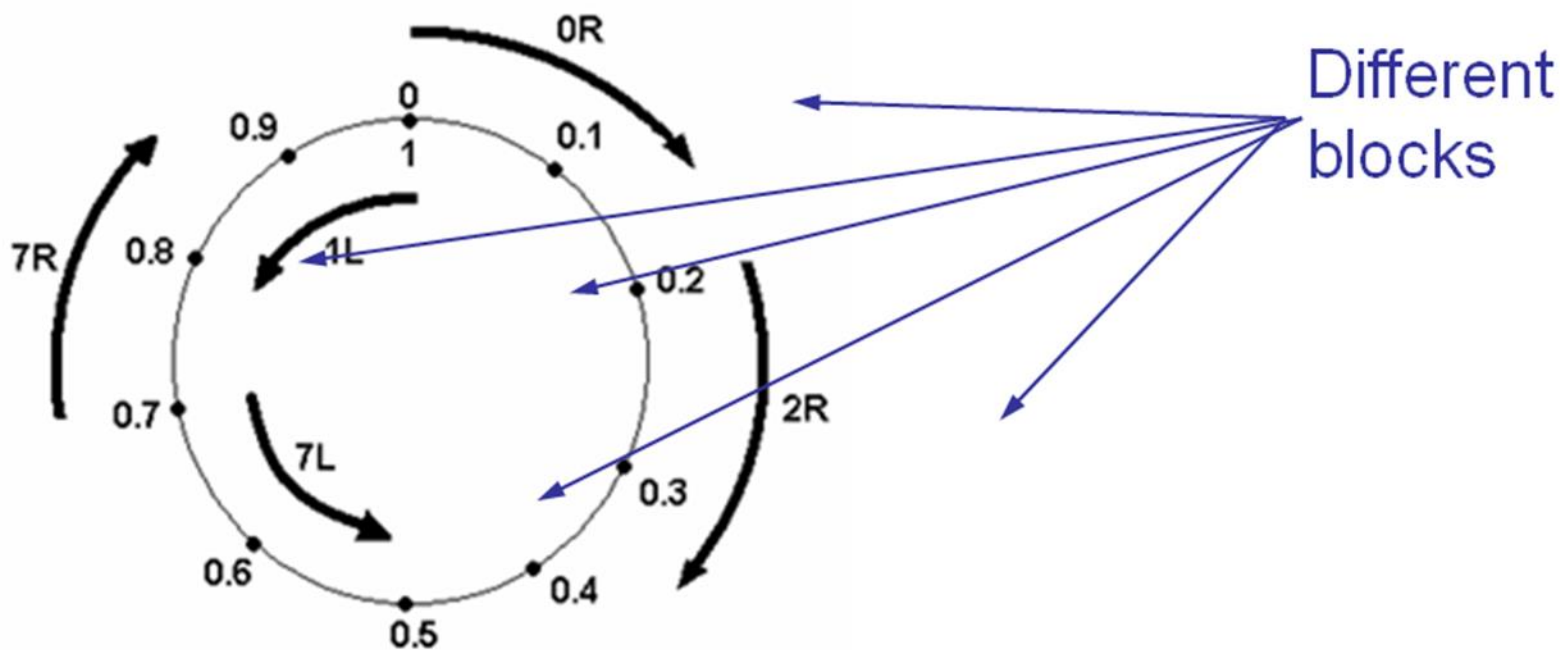
These sampling designs were selected because they are used at SCB (more or less common) and because they have fix sample size and are possible to combine with coordination by PRNs (except systematic sampling)

Software for sample selection

- Completely user supplied - a SAS-dataset including the frame population for the specific survey together with the sampling design and other necessary information for the sample selection must be provided.
- Other parameters needed depends on chosen sampling selection method.
- Output from the tool is a SAS-dataset: a copy of the frame population complemented with an indicator on whether the unit is included in the sample or not, inclusion probabilities and sampling weights.

Transformation of PRNs

For sequential SRS we choose the starting point and the direction, and include the necessary number of units:



Transformation of PRNs

The objective of the transformation is to select the n first units to the right of zero, independently of the chosen starting point and sampling direction.

Transform the PRN u_k into z_k as follows

Starting point S and direction right: $z_k = \text{Mod}(u_k - S + 1, 1)$

Starting point S and direction left: $z_k = \text{Mod}(u_k - S, 1)$

Note that the transformed z_k only is used in the sample selection phase, the original PRNs is not changed. This transformation is used for all sampling designs (necessary for Pareto π ps sampling)

Business structure in Sweden

Frozen version of the Swedish BR November 2018 (Legal Units)

Size class	Number	Number	Employees	TurnOver
Small: < 20	1 198 513	98,1	22,0	27,1
<i>thereof: 0</i>	883 544	72,3	0,0	8,0
<i>thereof: 1-19</i>	314 969	25,8	22,0	19,1
Medium: 20-199	20 719	1,7	21,6	26,7
Large: \geq 200	2 106	0,2	56,4	46,2
Total	1 221 338	100	100	100

Legal Units burdened from surveys in the SAMU during the year 2017

Number of surveys	Total	Total SAMU	Inactive SAMU	0 Empl. SAMU	1-4 Empl. SAMU	5-9 Empl. SAMU	10-19 Empl. SAMU	20-49 Empl. SAMU	50-99 Empl. SAMU	100- Empl. SAMU
1	97 217	22 950	365	7 761	7 149	3 186	3 074	1 224	187	4
2	27 209	17 603	4	1 465	6 137	3 965	3 414	2 118	464	36
3	10 456	9 412	1	638	1 908	1 786	2 423	2 003	509	144
4	5 001	4 823	0	172	406	529	1 383	1 577	538	218
5	3 266	3 243	0	55	77	141	761	1 313	526	370
6	2 104	2 100	0	15	12	56	361	869	411	376
7	1 363	1 363	0	9	12	18	139	515	354	316
8	1 055	1 055	0	2	1	0	65	383	335	269
9	708	708	0	3	1	1	31	198	225	249
10	574	574	0	1	0	2	20	134	178	239
11	440	440	0	0	1	0	4	73	138	224
12	389	389	0	0	0	1	3	40	101	244
13	296	296	0	0	0	0	0	18	84	194
14	248	248	0	0	0	0	0	13	53	182
>=15	1 215	1 215	0	0	0	0	0	8	64	1 143
At least 1	151 541	66 419	370	10 121	15 704	9 685	11 678	10 486	4 167	4 208
Total BR	1 218 041	1 218 041	1 654	851 934	263 879	49 447	26 764	15 452	4 686	4 225
Prop, %	12,4	5,5	22,4	1,2	6,0	19,6	43,6	67,9	88,9	99,6

The SAMU-system

- *Advantages:*
 - very easy to monitor and maintain
 - it can be shown that SAMU produces probability samples
- *Disadvantages:*
 - no exact control of overlap and rotation
 - uncontrolled jumps in and out of a sample due to changes in the business population or due to changes in the sampling design