Sampling Coordination for Business Surveys at Statistics Netherlands

Marc Smeets (mset@cbs.nl) and Harm Jan Boonstra (hbta@cbs.nl)

EESW17 - August 31, 2017

Introduction

- Aims of coordinated sampling system
  - Support sampling for (all) Dutch business surveys.
  - Even spread of total survey burden over enterprises, both over time and surveys.
- Current situation
  - Coordinated sampling for 17 Dutch business surveys.
  - Spread of survey burden over time for these surveys.
  - Spread of survey burden over surveys for Structural Business Survey and Investment Survey: survey-free periods up to 15 years for small-sized enterprises.
- Methodology of coordinated sampling system
  - Based on former EDS system (Huis et al., 1994).
  - Sampling algorithm implemented in R-package SBS.

Sampling algorithm

- Given group $\mathcal{G}$ of surveys with common sampling frame U.
- Both stratified cross-sectional and rotating panels can be combined in $\mathcal{G}$.
- Assign to every $k \in U$:
  - $R_k \in [0, 1]$: unique permanent random number (PRN).
  - $B_k \geq 0$: total built up survey burden in $\mathcal{G}$.
  - $I_{pk} \in \{0, 1\}$: panel membership for every panel $p \in \mathcal{G}$.
- For every stratum h draw sample of size n_h:
  1. Sort units by specified ordering of (R_k, B_k, I_{pk}).
  2. Select first n_h units.
  3. Update values of (R_k, B_k, I_{pk}).

Illustration of sampling algorithm

Figure 2. Enterprises in random order

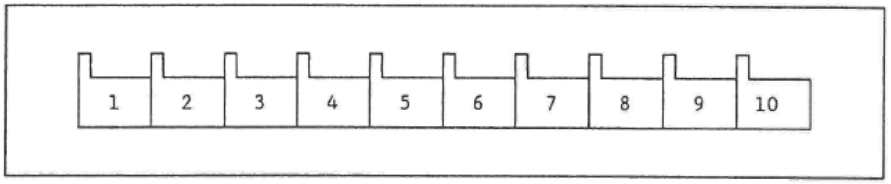


Figure 3. Before the second sample

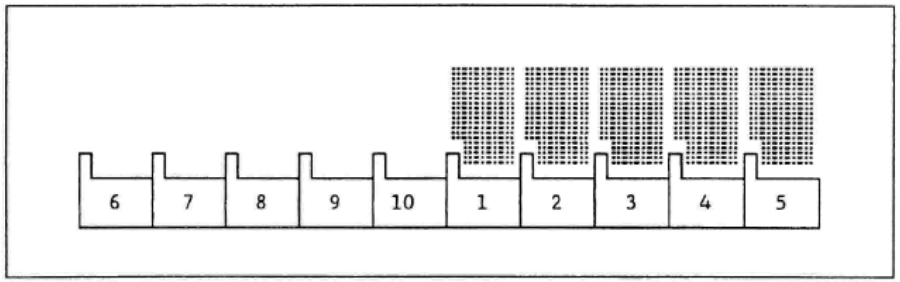


Figure 4. Before the third sample

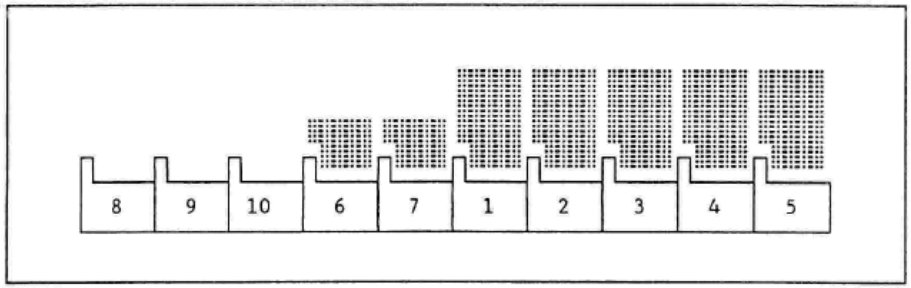
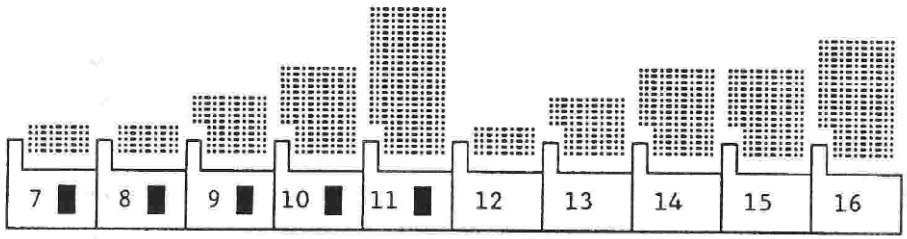


Illustration of panel rotation



Situations:

1. rotation fraction $v_h = 0.2$, sample size $n_h = 5$: 11 out, 12 in.
2. rotation fraction $v_h = 0.2$, sample size $n_h = 4$: 11 out.
3. rotation fraction $v_h = 0.2$, sample size $n_h = 3$: 10, 11 out.

Population dynamics

- Assign appropriate (R_k, B_k, I_{pk}) to births and stratum movers
  - before every draw of a sample in $\mathcal{G}$,
  - such that births, stratum movers and existing units in stratum h have same joint distribution of (R_k, B_k, I_{pk}).
- Births
  - assign new $R_k \in [0, 1]$,
  - copy (B_k, I_{pk}) from existing unit in h closest to R_k.
- Stratum movers
  - determine relative position of stratum mover in old stratum,
  - copy (B_k, I_{pk}) from existing unit in new stratum closest to relative position. A new R_k close to PRN of existing unit is assigned.
  - Possible orderings: (i) by R_k, (ii) by B_k,R_k or (iii) by I_{pk},B_k,R_k.

Basic and substratification

- Basic stratification: common stratification for surveys in $\mathcal{G}$.
- Depart from basic stratification possible by use of substrata.
  - Define parameters (R_k, B_k, I_{pk}) at basic stratum level.
  - Sampling is done per substratum.
  - Spread of survey burden is suboptimal.
- For cross-sectional surveys no restrictions.
- For panels:
  - Maximal 3 substrata in basic stratum h with fractions f_{h1}, 0, 1.
  - Panel indicator I_{pk} denotes imaginary panel.
  - Real panel can be derived from I_{pk}.

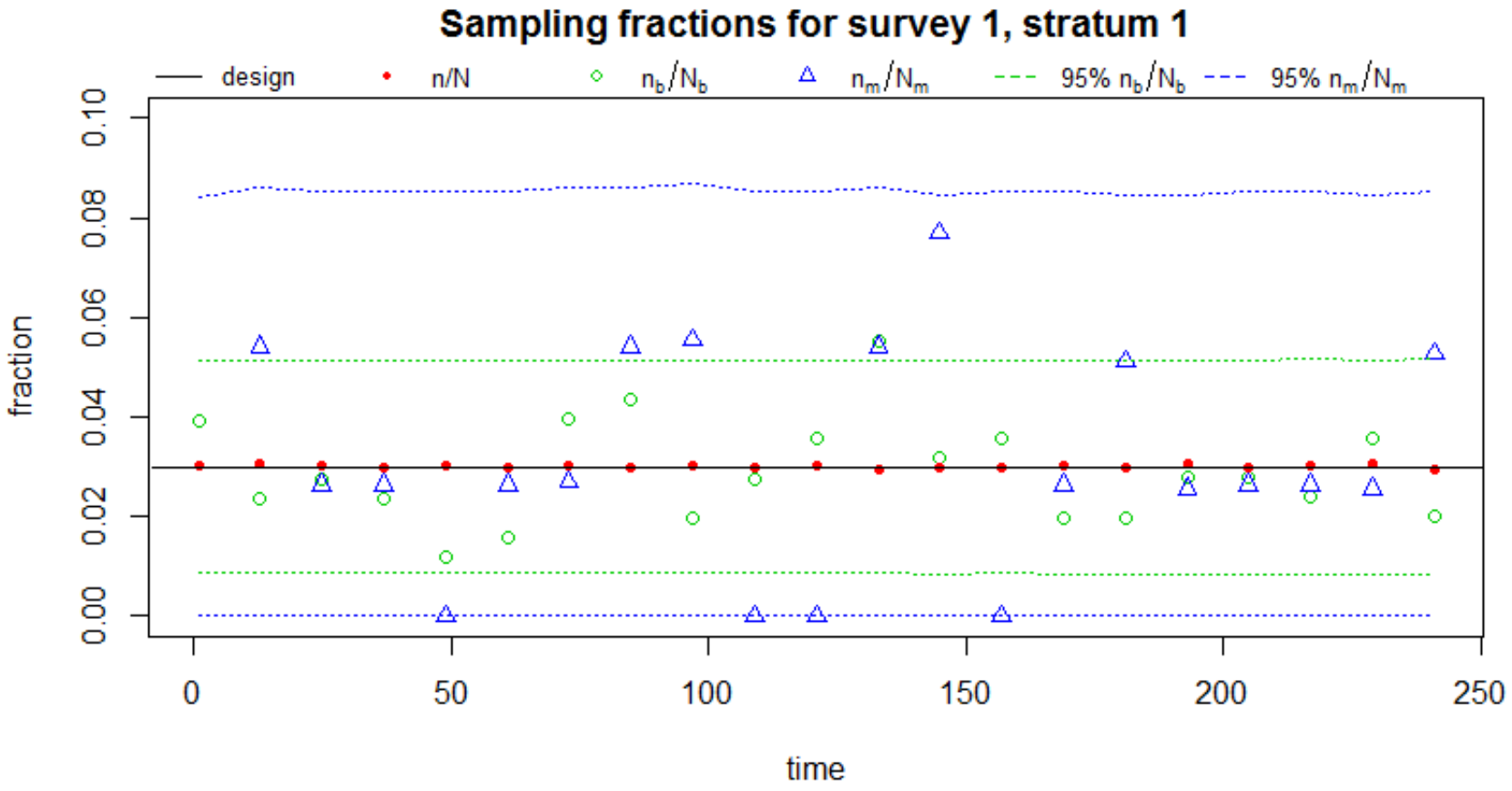
Simulation study

- Coordination of sampling in group of 3 surveys
  - by simulating a series of 250 monthly draws,
  - from population with 100,000 units, 5 basic strata and
  - simulated population dynamics.
- Surveys with sampling fractions:

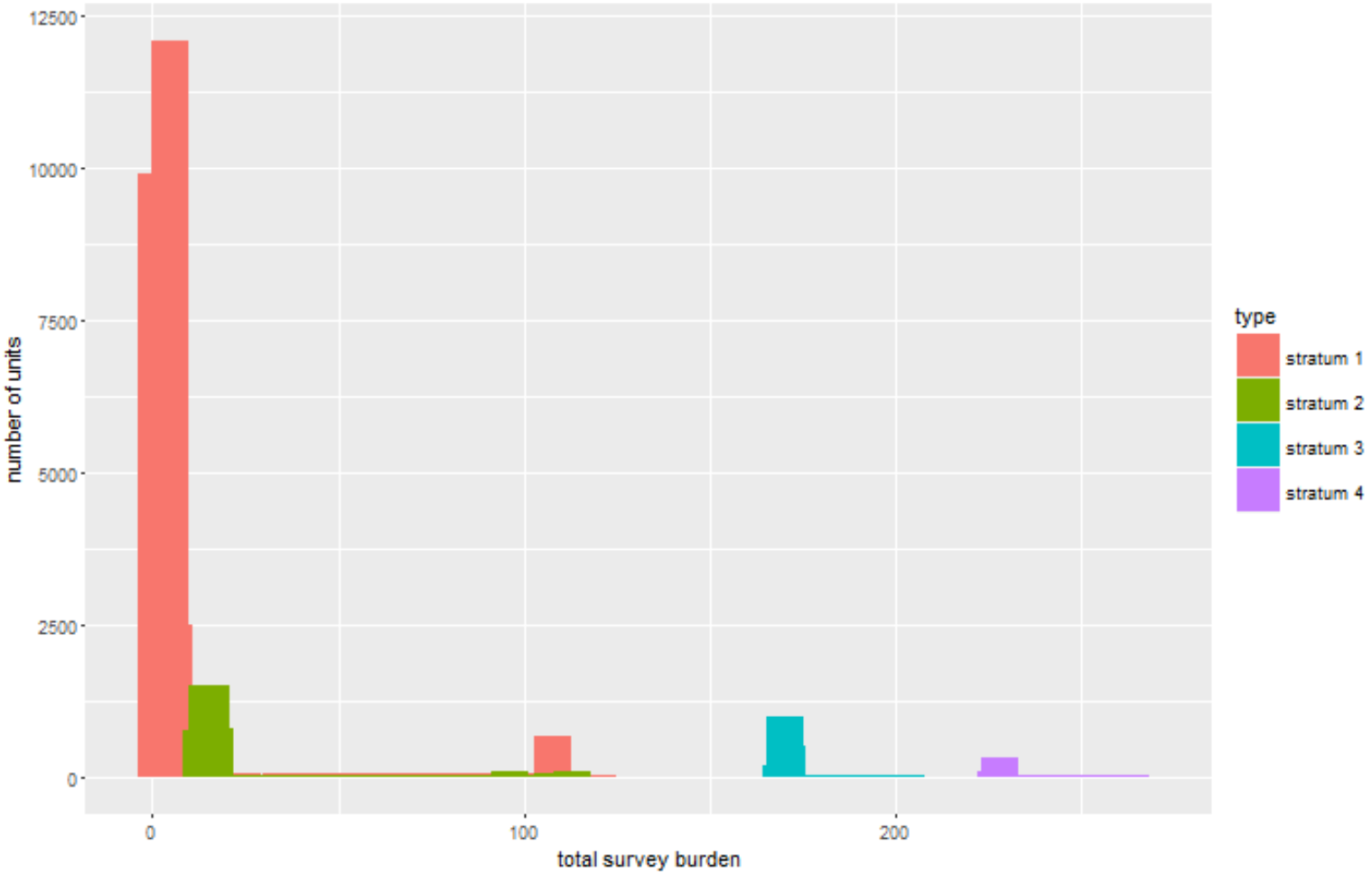
Survey	Frequency	Rotation	1	2	3	4	5
1 (no panel)	year	-	0.03	0.06	0.1	0.15	0.3
2 (panel)	month	0.1 (yearly)	0.02	0.06	0.1	0.15	0.3
3 (panel)	month	0.2 (monthly)	0.01	0.05	0.6	0.8	1

- Aspects of spread of survey burden:
  - survey-free periods,
  - length of stay in panel,
  - multiple draws in group at same time.

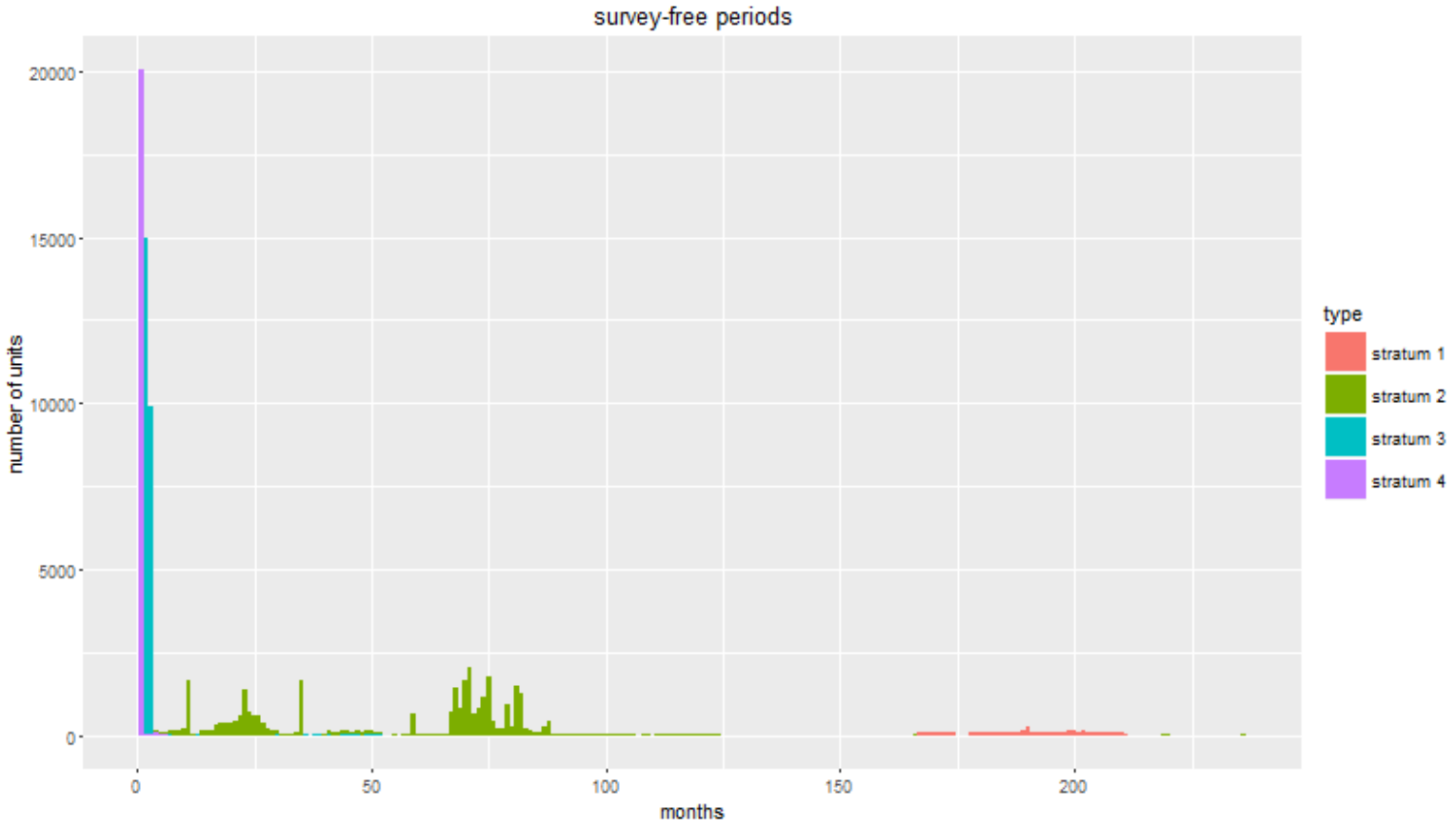
Realised sampling fractions



Total survey burden



Survey-free periods



Discussion

- Questions or comments?
- Improvement of handling stratum movers when more panels are combined in a group?
- Substratification for panels without restrictions?
- Extension of sampling coordination to more complex sampling designs or to larger groups of surveys?