Updating and monitoring the NACE code of single-establishment enterprises in the Swiss business register

WIESBADEN GROUP ON BUSINESS REGISTERS
24-27 September 2018, Neuchâtel

Monika Ferster

Swiss Federal Statistical Office
Statistical Methods Unit

Outline

Overview

- The NACE code is an important variable for all units in the Swiss business register (BR)
- Quality should be managed

- NACE code of single-establishment enterprises (S-ENT)
  - updated by other sources **OR**
  - regular control/update based on quarterly samples
    $\longrightarrow$ *main part of presentation*

- NACE code of multi-establishment enterprises (M-ENT; part of BR\S-ENT)
  - quarterly update (profiling)
  - yearly update (profiling light)

Overview - continuation

- Sampling probabilities
  - by enterprise size classes
  - should lead to an expected quarterly sample of 10'000 enterprises
  - should meet requirements on actualisation frequencies
- First frame in april 2014 (second quarter)
- Quarterly update of first frame
  - gives history of passed and future control for a single enterprise or a group of enterprises

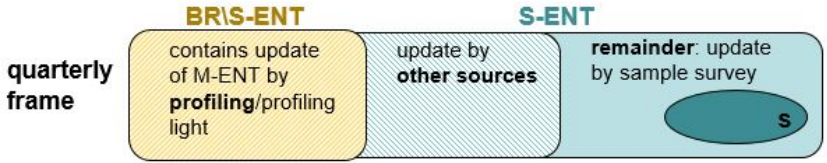
Frame and enterprise types

Each quarter an extract of the BR is taken:

- All active Swiss enterprises, divided by size classes

size class	very small	small	midsize	large
number of employees	1 - 2	3 - 9	10 - 99	$\geq$ 100

- Actualisation of NACE code by 3 types of control/update



Constraints to sample survey

- Quarterly sample size of about 10'000 enterprises
- Actualisation cycle of different lengths
  - two years for large enterprises
  - three years for midsize enterprises
  - four years for small enterprises
  - five years for very small enterprises
- Dynamic changes in the BR
  - enterprises can change structure, size or type
  - number of active enterprises is not stable over time (mostly increasing slightly)

Random numbers: idea

- To each enterprise of the frame belongs a permanent random number uniformly distributed in the interval $[0, 1]$.

  For each size class:
- The interval $[0, 1]$ is divided into subintervals whose lengths reflect sampling rates of different quarters, taking into account expected population increases in the BR.
- Each quarterly sample is defined by enterprises with random numbers contained in the corresponding **selection interval**.
- In other words we have a Poisson sampling scheme with permanent random numbers.
- The quarterly samples are negatively coordinated.

Random numbers: calculation of selection interval for each size class

m_d = length of selection interval of quarter d
geometric series $s_k = m_1 + m_2 + ... + m_k = m_1 \cdot \frac{1-q^k}{1-q}$ with $\frac{m_{d+1}}{m_d} = q = \frac{1}{rate\_of\_increase}$

A cycle is controlled in k quarters if $s_k \geq 1$, k minimal.

We define $m_1 = \frac{n_s}{N_{remainder}}$ with n_s the desired sample size.

Taking the average of the ratios $N_{remainder} : N_{S-ENT}$ of past five quarters, denoted by ratio in our calculation, we can have a more stable m_1 by $m_1 = \frac{n_s}{ratio \cdot N_{S-ENT}}$.
We fixed n_s in a way, that we obtained satisfactory cycle lengths.

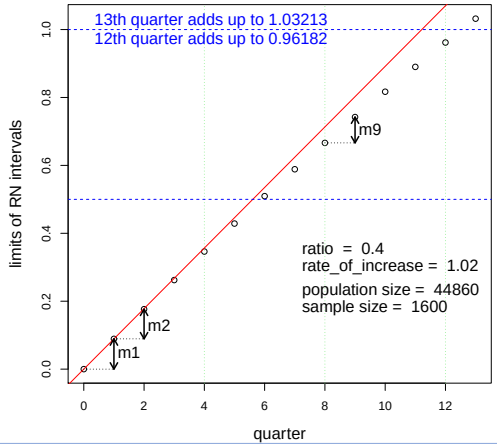
Random numbers: calculation of selection interval for each size class - continuation

Starting positions for the geometric series calculation for each size class

	very small	small	midsize	large
assumptions**				
ratio	0.36	0.41	0.40	0.25
rate_of_increase	1.02	1.005	1.02	1.01
given (first frame)				
population size N_{S-ENT}	332'502	141'965	44'860	2'098
final sample size n_s	4'700	3'600	1'600	100

**means of five past quarters

Random numbers: situation for the midsize class



Random numbers: extract of RN-table

Random number interval limits (line 2014_2 = m_1)

period	very small	small	midsize	large
2014_2	0.0393	0.0618	0.0892	0.1907
2014_3	0.0778	0.1234	0.1766	0.3794
2014_4	0.1155	0.1846	0.2623	0.5663
2015_1	0.1525	0.2456	0.3463	0.7514
2015_2	0.1888	0.3062	0.4287	0.9346
2015_3				1
2015_3	0.2243	0.3665	0.5094	0.1160
2015_4	0.2592	0.4265	0.5886	0.2956
2016_1	0.2934	0.4863	0.6662	0.4734
2016_2	0.3269	0.5457	0.7424	0.6495
2016_3	0.3598	0.6048	0.8170	0.8238
2016_4	0.3920	0.6637	0.8901	0.9964
2017_1				1
2017_1	0.4235	0.7222	0.9618	0.1673
2017_2			1	
2017_2	0.4545	0.7805	0.0321	0.3365
...				

rounded to four digits

Random numbers: result for first frame

S-ENT: Number of enterprises by size class and type of NACE code control type

quarter	control of NACE code by other sources	remainder	(in sample)	$N_{remainder} : N_{S-ENT}$
very small	212,716	119,786	(4,750)	36%
small	82,766	59,199	(3,711)	42%
midsize	27,115	17,745	(1,576)	40%
large	1,690	408	(84)	19%
total	324,287	197,138	(10,121)	38%

Trade-off between length of cycles and sample size
($\frac{119,786}{4,750} \approx 25$; more than 6 years)

Monitoring the NACE code control

The first frame is updated quarterly: **monitoring dataset (MD)**
The MD reflects always the current situation. It increases every time by about 20,000 enterprises.

Variables of the monitoring dataset

enterprise identification variables enterprise identifier and its permanent random number *status_rn*: since when continuously in MD or when *leaving*
current quarter structural variables like NACE code, size class, NACE code control type and if the enterprise is drawn in the sample
other quarters same information as for current quarter for the entering and the penultimate quarter

Monitoring dataset up to 2018_1 : overview

status_rn < 0 means that the enterprise does not belong to the current frame.

	status_rn	Frequency	Percent	CumFreq
no more active since	-2018_1	18,463	1.66	18,463
no more active since	-2017_4	17,357	1.56	35,820
no more active since	-2017_3	6,685	0.6	42,505
no more active since	-2017_2	32,638	2.94	75,143
no more active since	-2017_1	27,560	2.48	102,703
...	...	...	...	...
no more active since	-2014_4	11,961	1.08	210,293
no more active since	-2014_3	15,090	1.36	225,383
active since beginning	2014_2	626,949	56.5	852,332
constantly active since	2014_3	12,791	1.15	865,123
constantly active since	2014_4	12,833	1.16	877,956
...	...	...	...	...
constantly active since	2017_1	15,662	1.41	102,4816
constantly active since	2017_2	34,618	3.12	105,9434
constantly active since	2017_3	7,414	0.67	1,066,848
constantly active since	2017_4	14,597	1.32	1,081,445
constantly active since	2018_1	28,243	2.55	1,109,688

Monitoring dataset up to 2018_1 : Applications

The last updated MD includes all quarters from 2014_2 up to 2018_1.

Gaps

Of the 232,921 active remainder part enterprises in the quarter 2018_1 almost 0.4% had been intermediately inactive (their *status_rn* was negative during some quarters).

First sample

Of the first sample (10,121 enterprises in 2014_2) 1,833 are no more active, 65 were only inactive during some quarters in-between and 8,223 enterprises are active since the beginning in 2014_2.

Control of large enterprises of 2014_2

The starting (2014_2) remainder part had 408 large enterprises. For 2018_1 we see, that 53 enterprises had never been controlled by a sample. They changed their NACE code control type, their size class or had been intermediately inactive. At best they could have been controlled in another NACE code control type - obviously not in the remainder part.

Adaptations

- System change in BR hugely increased the number of very small enterprises
  $\Rightarrow$ very small enterprises are neglected from 2015 on if full-time equivalent < 0.5 (cut-off)

- In 2016 the set of enterprises for which NACE code update is obtained by other sources has been extended. This didn't change the whole process, only the criteria for the type of NACE code control had to be adapted.

Conclusions

- By the presented approach based on permanent random numbers we have a consistent managing system of the NACE code control for each enterprise.
- The system can easliy be adapted to changes of the population size (e.g. increasing number of enterprises) or in resources (manner of how controls are done).
- Quarterly sampling can be executed routinely.
- Structural changes of the enterprises can be noticed (e.g. changing from M-ENT to S-ENT).
- Past and future control of the NACE code is known (by the random number).

References

Ohlsson, E. (1995). Coordination of samples using permanent random numbers. In Cox, B. G. et al., editors, *Business Survey Methods*, chapter 9, pages 153-169. Wiley inc., New York, USA.

SÄRNDAL, Swensson B., Wretman J.: *Model Assisted Survey Sampling*. Springer-Verlag New York, 1992.

Thank you for your attention!