# A system for continuous control of the NACE code in the Swiss business register

Monika Ferster[1]

**Abstract**
The NACE code is an important variable for all units included in the Swiss business register. It describes the main economic activity of an establishment and is used in several statistical processes: sampling frames, sampling designs, domain extrapolation, etc. The NACE code of multi-establishment enterprises are controlled or updated by an ongoing process. For single-establishment enterprises, the control of NACE codes should be planned consistently to ensure that their NACE code in the register is updated in a systematic manner.

At the Swiss Federal Statistical Office we have developed and implemented a system to control the NACE code for these single-establishment enterprises. Every quarter a sample of establishments is selected to control the NACE code. The underlying sampling scheme is Poisson with sampling weights depending on establishments' sizes. There are two restrictions for the sampling design: expected quarterly sample size and length of the cycles in which all establishments of a size category should be controlled. For example, all the big establishments with at least one hundred employees should be controlled within two years. To handle the control process, we assign to each establishment a permanent random number uniformly distributed in [0;1]. Respecting the dynamic changes in the register, we split the [0;1]-interval for each size category into a number of selection intervals so that we can reach the demanded cycle length. Consequently, each quarterly sample will contain all the establishments of the random number subinterval belonging to this quarter and size category. Keeping the permanent random number of an establishment, we know when the NACE code of the establishment has been controlled for the last time or when the next control will take place.

**Key Words:** NACE code, single-establishment enterprises, permanent random number

## 1. Introduction

The NACE code describes the main economic activity of an enterprise. It is an important variable in a business register and should reflect the real situation of the enterprise as recently as possible. Former controls or changes of the NACE code should be listed in the history of the NACE code of the examined enterprise.

An up-to-date NACE code should be garanteed, especially for the large enterprises.There are different ways, by which the NACE code is controlled. Some enterprises are regularly controlled by an automatic procedure, but there is a remainder of mostly single establishment enterprises where the control must be planned. This takes place by quarterly samples, so that for example the NACE code of the large enterprises is controlled within two years and of the very small ones within five years.

The result of a control can be that there was an error in the codification, which has to be corrected, that there was really a change in the economic activity or that the control was just a control without change. In the following we describe how the system of control is handled: which enterprise will be controlled by which quarterly sample. We are not interested in the result of the control itself.

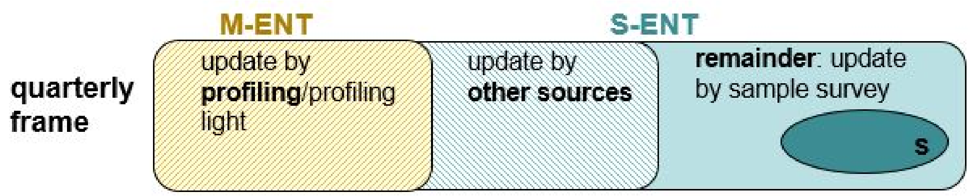
[1]Monika Ferster, Swiss Federal Statistical Office, Statistical Methods Unit, Espace de l'Europe 10, 2010 Neuchâtel, Switzerland, email: monika.ferster@bfs.admin.ch

## 2. Quarterly frame and NACE code control types

Each quarterly frame is built by all active Swiss enterprises in the Swiss business register. It is stratified by four size classes: large ($\geq 100$ employees), midsize ($10 - 99$ employees), small ($3 - 9$ employees) and very small (1 or 2 employees). The first frame dates from April 2014 (second quarter) and consisted of 801,734 enterprises.

There are three types of NACE code control or update. The NACE code of a multi-establishment enterprise (*M-ENT*) is controlled by the so called profiling or profiling light procedure. Enterprises in the profiling are regularly updated - quarterly for profiling and yearly for profiling light enterprises. For the single-establishment enterprises (*S-ENT*) there is either a forced control of the NACE code by different sources - for example when the enterprise is recently and for the first time included in the business register - or for all other enterprises (the *remainder*) by a quarterly sample survey.

**Figure 1**: Types of NACE code control




Remark: We are using the words *control* or *update* of the NACE code synonymously.

## 3. Constraints to the sample survey

There are two competing constraints to the sample survey. On the one hand our staff responsible for the NACE code can control the NACE code of about 10,000 enterprises in a quarter. On the other hand there is a requested maximum cycle length by which the NACE code of all enterprises of a size class should ideally be updated : After eight quarterly samples, which corresponds to two years, the economic activity of all large enterprises should have been controlled. The NACE code of midsize enterprises should have undergone an update after three years, of small ones after four and of very small after five years.

It should not be forgotten that the business register is changing continuously: an enterprise can change its structure (i.e. from S-ENT to M-ENT), its number of employees and hence the size class or the way its NACE code is controlled (e.g. it was part of the profiling procedure of the past quarter and is in the *remainder* of the actual quarter).

Furthermore the number of active enterprises in the business register is not stable over time but slightly increasing.

There is a trade off between sample size and cycle lengths. It leads to taking into account that the large, midsize and small enterprises are controlled within the demanded cycle length but expanding the cycle length of the very small enterprises so that the sample size of about 10,000 is maintained.

## 4. Random numbers

Starting with the first frame in April 2014 and ongoing with subsequent quarterly frames each enterprise entering this system receives a permanent random number uniformly distributed in the interval $[0, 1]$. The following procedure is applied to each stratum:

- The interval $[0, 1]$ is divided into subintervals whose lengths reflect sampling rates of different quarters and dynamic changes in the Swiss business register.
- Each quarterly sample is defined by enterprises with random numbers contained in the corresponding **selection interval**.
- The quarterly samples are therefore negatively coordinated.

So we have a Poisson sampling scheme with permanent random numbers.

### 4.1 Calculation of selection intervals for each stratum

If we suppose that the business register were static (no dynamic changes) then for example for the midsize stratum the $[0, 1]$ interval whould be divided into twelve subintervals of same length $\frac{1}{12}$. In the context of sampling this interval length of $\frac{1}{12}$ can be seen as a sampling rate. If we add up all selection intervals until they cover the whole interval $[0, 1]$, we will have reached the end of the midsize cycle because every enterprise (= random number) has been selected.

However, because the Swiss business register is increasing slightly, the random numbers become denser and therefore the length of the selection interval slightly shorter if we want to keep about the same number of enterprises (= random numbers) in a selection interval. The assumption of a constant rate of increase leads to a geometric series, where we are interested to know, when its sum is greater than 1.

With $m_d$ as length of the the selection interval of quarter $d$ we have the geometric series $s_k = m_1 + m_2 + ... + m_k = m_1 \cdot \frac{1-q^k}{1-q}$ with $\frac{m_{d+1}}{m_d} = q = \frac{1}{rate\_of\_increase} \forall d \in \{1, 2, ..., k\}$ and $m_1$ the first selection interval length or the first sampling rate. A cycle is controlled in $k$ quarters if $s_k \geq 1$, $k$ minimal.

To be able to calculate the geometric series we need values for $m_1$ and $q$. The latter is explained before and $m_1$ is the sampling rate for the first sample of quarter 2014_2 (April 2014). Using the notations of Figure 1 we could define $m_1 = \frac{n_s}{N_{remainder}}$ with $n_s$ the sample size of the first sample and $N_{remainder}$ the size of the remainder control type of the first frame. To be more stable we include the $ratio$ $N_{remainder} : N_{S\_ENT}$ of past quarters in our calculation. Table 1 shows for each stratum the values used for the calculation of the first quarter selection intervals $m_1$. We tried different sample sizes until we got to the final sample size $n_s$ with which we are close to our requested cycle lengths. Finally we calculate $m_1 = \frac{n_s}{ratio \cdot N_{S-ENT}}$.

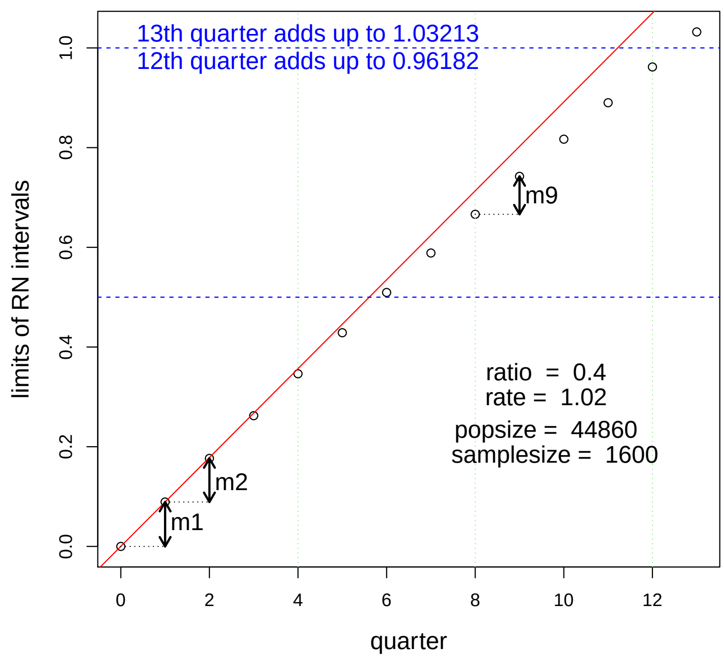
**Table 1**: Starting positions for the geometric series calculation for each stratum

| | 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 of April 2014) | | | | |
| 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

Figure 2 shows the different selection intervals $m_1, m_2, ...$ added up until they pass the unity. Thus the nineth selection interval $m_9$ signifies that the nineth midsize sample contains all midsize enterprises with random numbers in this interval. In Table 2 these two random number interval limits are the bold numbers of line 2016_1 and 2016_2.

**Figure 2**: random number interval limits for the midsize stratum




## 4.2 Extract of random number table

**Table 2**: Random number interval limits (line 2014_2 $= m_1$)

| # of quarter | quarter | very small | small | midsize | large |
|---|---|---|---|---|---|
| 1 | 2014_2 | 0.039264593 | 0.061848794 | 0.089166295 | 0.190657769 |
| 2 | 2014_3 | 0.077759291 | 0.123389883 | 0.176584232 | 0.379427838 |
| 3 | 2014_4 | 0.115499192 | 0.184624798 | 0.262288091 | 0.566328896 |
| 4 | 2015_1 | 0.152499094 | 0.245555061 | 0.346311482 | 0.751379448 |
| 5 | 2015_2 | 0.188773508 | 0.306182189 | 0.428687356 | 0.934597817 |
| 6 | 2015_3 | | | | **1** |
| 6 | 2015_3 | 0.224336660 | 0.366507689 | 0.509448017 | 0.116002143 |
| 7 | 2015_4 | 0.259202494 | 0.426533062 | 0.588625135 | 0.295610386 |
| 8 | 2016_1 | 0.293384685 | 0.486259801 | **0.666249761** | 0.473440330 |
| 9 | 2016_2 | 0.326896637 | 0.545689393 | **0.742352335** | 0.649509581 |
| 10 | 2016_3 | 0.359751491 | 0.604823315 | 0.816962702 | 0.823835572 |
| 11 | 2016_4 | 0.391962133 | 0.663663038 | 0.890110121 | 0.996435564 |
| 12 | 2017_1 | | | | **1** |
| 12 | 2017_1 | 0.423541194 | 0.722210026 | 0.961823276 | 0.167326644 |
| 13 | 2017_2 | | | **1** | |
| 13 | 2017_2 | 0.454501057 | 0.780465736 | 0.032130292 | 0.336525734 |
| ... | | | | | |

A unity in Table 2 signifies that in the corresponding quarter the control cycle will finish and a new one will start. That means all enterprises of this size class have been updated. All midsize enterprises will be controlled e.g. in the second quarter 2017, after precisely three years. Or the unity of the small ones lays in 2018_2 (not in Table 2).

### 4.3 Result for the first frame

Table 3 shows that the sampling fraction $\frac{4750}{115036}$ of the stratum of the very small enterprises is much smaller than the ones in the other strata: the very small stratum serves as buffer to fulfill cycle length requirements of large, midsize and small enterprises.

**Table 3**: Number of enterprises by strata and type of NACE code control type

| | control of NACE code by | | | | |
|---|---|---|---|---|---|
| size class | profiling (light) | other sources | remainder | (in sample) | total |
| very small | 248,165 | 212,716 | 119,786 | (4,750) | 580,667 |
| small | 15,336 | 82,766 | 59,199 | (3,711) | 157,301 |
| midsize | 13,219 | 27,115 | 17,745 | (1,576) | 58,079 |
| large | 3,582 | 1,690 | 408 | (84) | 5,680 |
| total | 280,302 | 324,287 | 197,138 | (10,121) | 801,727 |

## 5. Monitoring the NACE code control

The first frame of April 2014 is updated and completed by each quarterly frame. This leads to the monitoring dataset which allows to look at past and future NACE code control history.

### 5.1 Updating the first frame by quarterly frames

The monitoring dataset consists of three types of variables: enterprise identifying variables, status-variables which characterize the enterprise at the entering and at the current time-point and structural variables of each quarter. Each line of this monitoring dataset corresponds to an unique enterprise.

**Table 4**: Variables of the NACE code control monitoring dataset (MS)

| variable | description | handling |
|---|---|---|
| ent-Id | enterprise identificator | unique, can be overwritten by a new ent-Id |
| RN | permanent random number | unique, unchangeable |
| status_RN | currentness of the random number | year and quarter, when the random number entered or left the monitoring dataset (MS) |
| NACE_in | 6-digit NACE code | unchangeable, time-point when entering MS |
| emptot_in | number of employees | unchangeable, time-point when entering MS |
| NACE_cur | 6-digit NACE code | current NACE code of enterprise |
| emptot_cur | number of employees | current number of employees of an enterprise |
| scl_142 | size class (time-point 2014_2) | in the first frame i.e. second quarter 2014 |
| filter_142 | NACE code control type | in the first frame i.e. second quarter 2014 |
| brut_142 | if drawn in sample or not | 1 or 0 or missing (not in quarterly frame) |
| ... | | |
| scl_yyq | size class (time-point 20yy_q) | in quarterly frame of year yy and quarter q |
| filter_yyq | NACE code control type | in quarterly frame of year yy and quarter q |
| brut_yyq | if drawn in sample or not | in quarterly frame of year yy and quarter q |
| mutyyq_from | ent-Id before mutation | ent-Id of past quarterly frame |

The last four structural variables ..._yyq refer always to the current quarter. They are appended to the monitoring dataset at the end of each quarterly sampling procedure.
In the course of time the monitoring dataset can become very huge. Thus compressing and reducing variables of past quarters will be advisable.

### 5.2 Keeping or creating a random number

Each quarterly frame is based on an extract of the Swiss business register. By comparing it with the former frame it can be decided if

1. an ent-Id was also in the former frame $\longrightarrow$ the random number is kept
2. an ent-Id was not in the former frame $\longrightarrow$ a new random number is created
3. an ent-Id has changed $\longrightarrow$ the random number of the former ent-Id is taken.

Case 1 is by far the most frequent case, case 2 has a speciality - the gaps - and because of case 3 the mutation-variable *mutyyq_from* is necessary.
Concerning the gaps: Maybe an ent-Id was not in the former frame and is treated as a new ent-Id. But because "being an active enterprise in the Swiss business register" is sometimes volatile it can occur that this ent-Id was in a quarterly frame before the former one. Then it is not really a new enterprise and should not lose its NACE code control history.
If an ent-Id of the former quarterly frame is not found in the current frame and is not identified as the predecessor of a current ent-Id then this ent-Id disappears. To not lose this information its random number status receives the negative value of the current quarter to mark the time-point of the end of its NACE code control.

### 5.3 Evaluating the monitoring dataset

Table 5 shows the sizes of the different NACE codecontrol types and of the sample size for each quarterly frame.

**Table 5**: Number of enterprises in quarterly frames

| | control of NACE code by | | | | | |
|---|---|---|---|---|---|---|
| quarter | profiling (light) | other sources | remainder | (in sample) | total | ratio |
| 2014_2 | 280,303 | 324,288 | 197,138 | (10,121) | 801,729 | 38% |
| 2014_3 | 200,376 | 374,197 | 230,633 | (11,511) | 805,206 | 38% |
| 2014_4 | 193,223 | 386,768 | 230,523 | (10,337) | 810,514 | 37% |
| 2015_1 | 195,916 | 375,315 | 210,315 | (10,079) | 825,645 | 36% |
| 2015_2 | 198,551 | 377,172 | 214,662 | (10,057) | 836,134 | 36% |
| 2015_3 | 201,824 | 388,543 | 201,953 | (10,049) | 834,066 | 34% |
| 2015_4 | 205,880 | 386,477 | 206,655 | (10,218) | 844,896 | 35% |

The $ratio$ in Table 5 reflects for each quarter the proportion of the number of the enterprises of the $remainder$ in the population of all single establishment enterprises. It has the same meaning as the $ratio$ defined in chapter 4.1.
After the first three quarters a system change in the Swiss business register led to a huge increase in the number of mostly very small enterprises. To fulfill the demanded cycle lengths without increasing the sample size of about 10,000 units a cut-off had to be introduced: all enterprises with less than $0.5$ full-time equivalent were eliminated from the survey sampling process. After this adaptation a sample size of about 10,000 units could

be maintained as well as the random number interval limits of Table 2.

At the current quarter 2015_4 there are 926,433 enterprises in the monitoring dataset, 844,896 (91%) as active enterprises and 81,537 (9%) as inactive enterprises. *Inactive* signifies that during one or more quarters the enterprise was in the monitoring dataset as an active enterprise , but became inactive afterwards. Of all 844,896 active enterprises 730,609 (86%) entered the monitoring dataset at the beginning with the first frame in April 2014.

## 6. Adaptations

A system change in the Swiss business register in 2015 hugely increased the number of very small enterprises so that the very small enterprises with full-time equivalent $< 0.5$ were to be neglected from 2015 on. Without this cut-off the cycle length of the very small enterprises would become too long.

In 2016 another criteria for NACE code control by other sources had to be included. This didn't change the whole process, only the criteria for the type of NACE code control had to be adapted.

The random number table can easily be modified when changes happen in population (e.g. increasing of business register) or in resources (number of persons who control).

## 7. Conclusions

By the presented approach based on permanent random numbers we have a consistent managing system of the NACE code control for each enterprise in the Swiss business register. It is flexible to adapt to changing conditions as mentioned in 'Adaptations'. Quarterly sampling can be executed routinely.

With the monitoring dataset structural changes of the enterprises can be recognized. Past and future control of the NACE code is known.

## REFERENCES

Ohlsson, E. (1995) *"Coordination of Samples Using Permanent Random Numbers"*, in Business Survey Methods (eds B. G. Cox, D. A. Binder, B. N. Chinnappa, A. Christianson, M. J. Colledge and P. S. Kott), chapter 9, pages 153-169. John Wiley & Sons, Inc., Hoboken, NJ, USA.

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