7 th WORKSHOP ON LABOUR FORCE SURVEY METHODOLOGY

DATA PROCESSING AND DATA QUALITY

Madrid, Spain, 10 – 11 May 2012

A. Data processing: Processing of data with panel design and multi-mode data collection. Quality Controls and checks of data

A2–Plausibility checks and data validation for the Swiss LFS

Daniel Lörch - Switzerland

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

SLFS – Swiss Labour Force Survey

Validation of data from the Swiss Labour Force Survey (SLFS)

Contents

1 The SLFS

1.1 Chronology

The first SLFS was carried out in 1991. Until 2009 the survey was only held in the second quarter. Since 2003, alongside the normal random sample from the telephone directory, an additional sample has been taken from the register of foreigners. This ensures that the proportion of foreign people meets the requirements of the analysis. Since 2010 the SLFS has been carried out every quarter. Also in 2010, the Swiss Federal Council introduced the obligation to provide information. This decision was reversed in 2012 and the SLFS will once again be voluntary, probably from the 3rd quarter of 2012 onwards.

1.2 Overview of the methodology

The SLFS takes the form of a personal survey. It is designed as a rotating panel in which around 30% of the surveys carried out during each quarter are first surveys and 70% are follow-up surveys. On the basis of the 2-(2)-2 pattern, each person is surveyed four times over a period of 15 months.

The random sample for the SLFS consists primarily of the following two parts:

Approximately 80% from the telephone directory (Swiss and foreign resident population) → Two-stage sampling process: first the household and then the target person in the household.
Approximately 20% from the Register of Foreigners (Foreign resident population) → Single-stage sampling process: the target person is taken directly from the register.

(Status: Q2 2012)

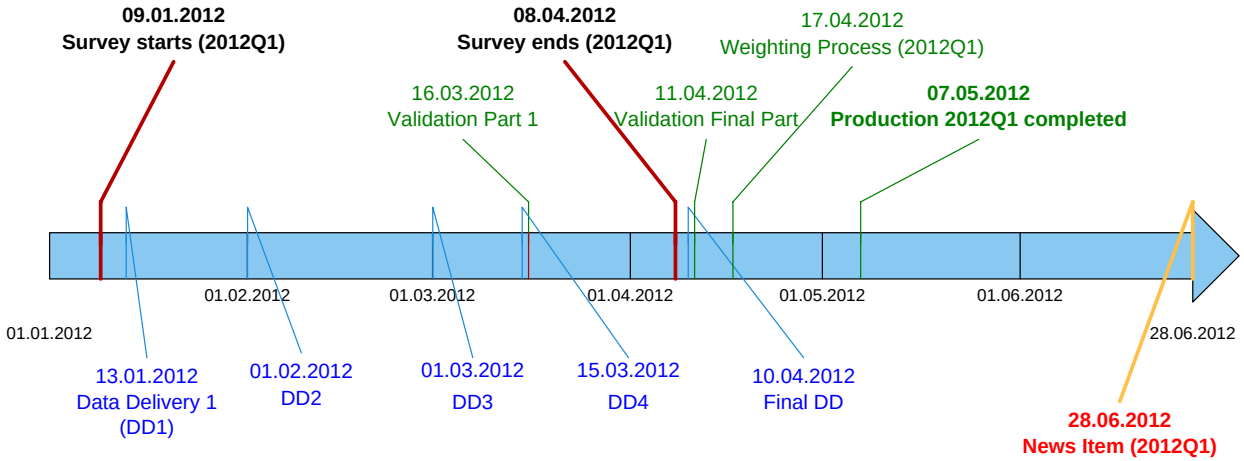
The survey is carried out only using CATIs (computer-assisted telephone interviews).

1.3 Production process

In order to meet the requirement for up-to-date information, the production margins for the SLFS are very tight.

For this reason, the production processes are brought forward as far as possible and are not delayed until the end of the survey.

Schedule for Data Preparation (SLFS 2012T1)



2 Validation phases

The SLFS production process consists of a number of validation phases with different objectives.

2.1 Before the survey

Before the survey, any changes to the questionnaire are validated. This validation phase is done on the one hand with online tests using interview simulations and on the other, with test interviews.

2.2 During the survey

2.2.1 Validation of the structure

One week after the start of the survey, frequency tables are used to evaluate the most important filters. These are simple cross-classified tables which show the proportion of interviewees per question. In particular, this enables nodes that must be passed by 100% of the participants to be evaluated effectively.

Question no.	Wave	Current Quarter		Last Year's Quarter		Difference
		abs	%	abs	%	%
Question 721.01	W1	960	100	1904	100	0
	W2	0	0	0	0	0
	W3	931	100	1431	100	0
	W4	0	0	0	0	0
Question 721.10	W1	425	44.3	859	45.1	-0.8
	W2	0	0	0	0	0
	W3	383	41.1	603	42.1	-1
	W4	0	0	0	0	0

The complexity of the questionnaire makes systematic validation of the filters difficult.

2.2.2 Online validation of combinations of variables

The combination of the “working hours per week” and “workload” variables is used as an illustration. If the full-time position exceeds the predefined limits, the interviewer is shown a visual warning.

2.2.3 Online validation of variable characteristics

All the data entered is validated using programmed data checks. Data outside the validity framework is not accepted. For example, the gender variable “sex” can only be given the codes 1 or 2. Birth dates must be entered in a valid 6-digit date format.

2.2.4 Online validation of variations over time

In addition to the selective validation of variable characteristics, longitudinal validations are carried out of the following socio-demographic variables: gender (SEX), birth date (YEARBIR, DATEBIR), marital status (MARSTAT), nationality (NATIONAL) and residence status.
If the current answer differs logically from the answer given in the previous survey, the current answer is repeated in a confirmation question and can be corrected, if necessary.

2.3 After the survey

After the survey, the socio-demographic and income variables are validated. As register data is only available for people from the foreign resident population taken from the register (approximately 20% of the sample), it is important to ask why variables are being validated when:

a) there is no reliable source of corrections for the majority of the variables.
b) limited resources are available.
c) there is a risk of bias.

The subsequent discussion leads to the following practical conclusions:

i) Variables with a register reference should be validated to ensure consistency with the register and over time.
ii) Variables which are used in the weighting model should be validated to ensure consistency over time.
iii) Variables with characteristics which are obviously implausible should be validated to allow outliers to be corrected.
iv) The validation process should be automated as far as possible to improve efficiency.
v) As few variables as possible and as many as necessary should be validated for the sake of efficiency and stringency.

Conclusions i), ii) and v) result in the following variable sets being validated:

- Gender of the target person
- Date of birth of the target person
- Marital status of the target person
- Nationality of the target person
- Residence status of the target person

Conclusion iii) results in the variable “income from employment of the target person” being validated.

Variable combinations such as “birth date x year of highest educational qualification” are not validated. Although this example could be monitored using a simple rule (limit on the difference between the two dates), in the case of other combinations numerous difficulties are often involved.

2.3.1 Validation of socio-demographic variables

As mentioned in the previous section, only the variables which are important for the weighting model are validated:

- Gender
- Date of birth
- Marital status
- Nationality
- Residence status

a) Rules

The validation process is based on predefined rules.

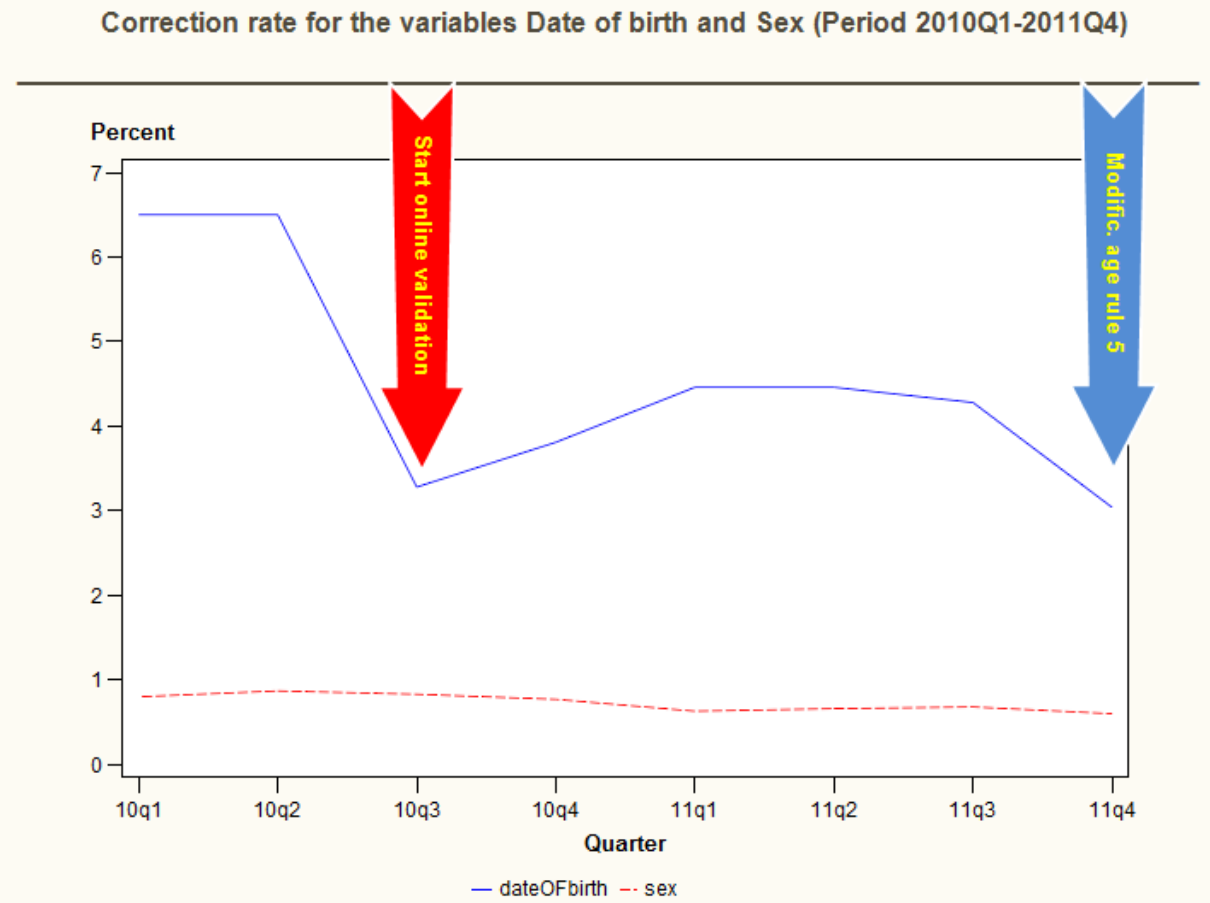
b) Corrections

Depending on the rule, corrections are made automatically or manually. Both the automatic and the manual processes are based on the Statistical Analysis System (SAS).

c) Flags

All corrections are marked with a flag so that they can be identified as corrections at any time.

d) Quantifying corrections



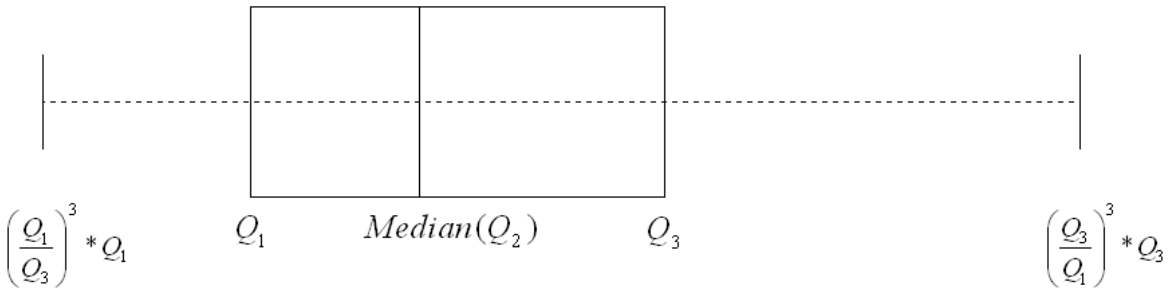
The large differences in the proportion of corrections for the “gender” and “birth date” variables are obvious. These are due largely to the difference in the complexity of the coding for the two variables.

The consequences of changes to the survey (online validation) or to the rules for the proportion of corrections are also clear. After a significant drop (-3.2 percentage points) as a result of online validation, the level for the birth date rose again, but still remained well below the previous figure (without online validation) (-2 percentage points).

2.3.2 Validation of income from employment

During the course of a survey, income is covered twice, in the first and third waves.

Depending on the type of income data given (gross/net, per hour/month/year), an upper and a lower limit is calculated using the following equation:



Figures outside these bandwidths are exported to an Excel file for manual post-processing using a variety of auxiliary variables (past figures, employment status, workload, median income from employment for specific professions etc.).

Following the validation process, the income from employment is annualised and finally converted into net or gross income.

3 Future developments

From 2013 onwards there will a complete register-based framework for sampling. This change will influence the methods currently used in the validation process.