Cross sectional weighting procedure

(Ref old QR : item 2.1.8; 2.1.8.1 2.1.8.2; 2.1.8.3; 2.1.8.41 (Reg. 28/2004 for IQR))

1. Design factor

The design effect term is the ratio of variance of a certain statistics under the actual design to that variance under a simple random sample of the same size.

$Deft^2 = 1.25$.

2. Non-response adjustments

The non-response adjustments are introduced to reduce the effect of differences in response rates achieved in different parts of the sample.

A description on how the design factor is corrected (the area and household level characteristics) taken into account the differences in response rates achieved in different parts of the sample will be provided.

Cross-sectional data

Non-response adjustment

The household design weights (**DB080**, see below) were multiplied by $n_{sample,h} / n_{respondents,h}$ in every stratum ***h*** *(*inverse of the response probability in every stratum) in the survey year.

Design weights

Separately calculated to each rotational group from the master sample in the initial survey year (each of size 50 000 in the survey year 2012, 2011, 2010 and 2009) we had population figures for the person selection, i.e. where $\pi_{a,person\,k}$ is the inclusion probability of selection of eligible person ***k*** in the master sample. The inclusion probability of selection of eligible person in the household dwelling-unit (HH) he/she occupied in the master sample were $\pi_{ak} = \pi_{a,person\,k} n_{16+,HH\ of\ k}$. Using stratification (13 strata) and strata allocation for the household-dwelling unit at the second phase of the two-phase sampling (the effect of selecting the master sample and the SILC sample) the inclusion probabilities, calculated separately for each rotational group, were:

$\pi_k^* = \pi_{ak}\pi_{k|s_a}$ where $\pi_{ak} = \pi_{a,person\,k} n_{16+,HH\ of\ k} = n_{s_a} n_{16+,HH\ of\ k} / N$

and $\pi_{k|s_a} = n_h / N_{h,s_a}$ is a conditional probability at the second phase taking the stratification of the master sample into account.

The Dfile includes the design weight DB080 to the first year in the survey. **PB070** (personal design weight for selected respondent) of the Pfile is an inverse probability of the person (**DB080*$n_{16,HH}$**) calculated only to the household that had been accepted to the Pfile.

3. Adjustments to external data (level, variables used and sources)

A description of:

- The level (household, persons) of adjustment (calibration)
- The variables used in the adjustment
- Method applied will be provided

Cross-sectional data

The calibration can be interpreted as integrative, i.e. both the household and person levels were included in the process. The levels and variables included in the calibration were as follows:

1) Households: Region (≈NUTS-3), type of municipality by degree of urbanisation (national classification), household size, total sums of 15 different income variables.

2) Persons: Sex and age classes (0-4, 5-9, ... , 80-84, 85-).

The calibration variables were:
- Region (NUTS-3 level), capital region and Helsinki separated
- Size of household-dwelling unit (1,2,3,4,5,6,7,8+)
- Degree of urbanisation (designed by Statistics Finland)
- Men 0-4, 5-9, 10-14, ..., 80-84,85-
- Women 0-4, 5-9, 10-14, ..., 80-84,85-
- Income 1: Cash or near cash employee income
- Income 2: Income 1 > 0
- Income 3: Pensions
- Income 4: Unemployment benefits 1
- Income 5: Unemployment benefits 2
- Income 6: Income 4 > 0
- Income 7: Income from self-employment
- Income 8: Capital Income 1
- Income 9: Income from agriculture
- Income 10: Income from property and forestry
- Income 11: Other capital income
- Income 12: Income from forestry 2
- Income 13: Capital gains
- Income 14: Pensions > 0
- Income 15: Mortgage interests

Calibrations were performed by the raking method with the SAS supporting CALMAR macro.

Final cross-sectional weights

A description of:

- How cross-sectional weights are calculated will be provided

The non-response adjusted weights were scaled first for the fixed household size (housekeeping unit, 2,570,999 in the survey year 2012 process) of the survey year in each rotational group and then used as input weights for calibration of the accepted households, also in each rotational group separately. The weights were calibrated first to the distributions of household-dwelling units by region, degree of urbanisation, size, and of non-institutionalised household-dwelling population by sex*age, i.e. in order to get provisional weights. Later, when final income data from the total income distribution data of Statistics Finland were available, the calibrations of non-response adjusted weights were performed again with all calibration variables. The resulting weights for the accepted data units provide exactly the same margins as in the total income distribution data. To adjust the effect of different rotational groups for **DB090**, these weights were yet multiplied by the proportions of the accepted households of the rotational group (rg) of all accepted sample households:
$\mathbf{(}n_{respondents,\ rg}\ \mathbf{/}\ n_{respondents}\mathbf{)}$.

When DB090 is connected to the R file ("All persons currently living in households or temporarily absent") the **RB050** weights give the sum, which coincides with the exact size of the non-institutionalised household-dwelling population at the end of 2012, i.e. 5,319,813. Furthermore, when DB090 is linked to the P file ("All eligible persons for whom the information could be completed"), these weights **PB040** give the size of the population aged 16 years and over, i.e. 4,377,501. And linking DB090 to the sample persons in the Rfile or Pfile gives the size of the households defined (2,570,999). The variables values **DB090 = RB050 = PB040**.

Finally, the personal cross-sectional weight for the selected respondent, i.e. **PB060** is DB090 multiplied by $\mathbf{n_{16+,HH}}$**.** The size of the persons aged 16 and over is fixed (4,377,501).

The weight for the children aged 0 to 12, i.e. **RL070** ("children cross-sectional weight for child care) is calculated by multiplying RB050 with the term "number of non-institutionalised (household-dwelling population) children in age class X from registers"/ " number of children in age class X estimated with RB050", where X = 0 to 12.

Supplement sample of cross-sectional data

The weight processing of the supplement sample of the persons aged 16 selected to the second and the consecutive waves of the survey was integrated with the prime sample according to rotational groups from the survey year's non-response adjustment onwards. When design weights were calculated to the initial sample of the selection year, the DB080 values of the supplement units (the cross-sectional data contain DB080 only to the sample of the 1st wave) are missing in the data.