

Annex 5 – Weighting procedure

Design factor

Design weights are defined at the dwelling level and are equal to the inverse of the probability of inclusion that results from the survey design. Households and individuals have the same weight as the dwelling to which they belong.

Non-response adjustments

The *Statistiques sur les ressources et conditions de vie* (Statistics on income and living conditions, SILC) survey comprises 4 sub-samples: from incoming households (wave 1) to outgoing households being surveyed for the fourth and final time (wave 4). The latter entered the panel in year $n-3$. Under certain conditions, households entering in year n will be resurveyed until year $n+3$.

In order to correct for non-response within our panel, we are creating homogeneous response groups so as to calculate two specific probabilities:

- a) The probability of repeat collection in $n+1$
- b) The probability of response in n knowing that there will be repeat collection in $n+1$ *

(* in the specific context of individuals entering in n who leave the panel and are therefore not subject to repeat collection, we speak of probability of response in $n+3$, knowing that the individual's response status is not unknown)

Creation of homogeneous response groups

For households entering in n (1st wave), the available information is that contained within the sample or obtained by means of matching with other sources (in particular Filosofi $n-2$). The homogeneous response groups of incoming households are therefore modelled on the basis of exogenous data (at eh household level). The discriminant variables used are: the indicator for living in Paris, the income decile, the household type, age and occupancy status of the dwelling.

For households undergoing repeat surveys (2nd, 3rd and 4th waves), the homogeneous response groups are established at the individual level for each individual panel.

The discriminant variables used to model the homogeneous response groups are: socio-professional category; education; household type; dwelling type; neighbourhood type; age; gender; work situation; income decile; age and occupancy status of the dwelling; the indicator for severe material deprivation during the past year; the indicator for material deprivation during the past year; the indicator for the level of income poverty at the threshold of 60% of the median standard of living during the past year.

Each of these homogeneous response groups is assigned a response probability and a probability of repeat collection as the total number of households or individuals concerned (respondents or repeat collections, as applicable) within the response group over the total number of households or individuals within the group.

Link calculation and weight sharing

A link is assigned to all individuals present in n panels or non-panels, since the link between all individuals comprising a household (regardless of their status) is used for weight sharing.

An individual's link is the number of the last four years spent within the coverage of the survey.

A household's link is the sum of the links of the individuals of which it is comprised, regardless of their age or status (panel or not).

Weight sharing is a step specific to cross-sectional weighting that allows a weight to be recalculated for households that have seen the addition of one or more new members following the selection of the sample. Indeed, the weight of each household must be allocated such that the sample remains representative of the cross-sectional population.

Cross-sectional weighting

For panel individuals and based on the calculated probabilities of response and repeat collection, we define a cross-sectional probability of response as the product of the probability of repeat collection in n and the probability of response in n knowing that there will be repeat collection in $n+1$.

Let us take the example of an individual entering in $n-2$ who is therefore in the 3rd wave in n :

P_RCi is the probability of repeat collection in year i and P_REPi the probability of response in year i knowing that there will be repeat collection in year $i+1$.

In n , the individual's probability of response will be: $P_RC_{n-1} * P_RC_n * P_REP_n$

The household's probability of response is equal to the sum of the probabilities of response of the panel individuals of which it is comprised, divided by the household link.

The pre-calibration cross-sectional weight of the household is, ultimately, the drawing weight of the household divided by its probability of response.

Marginal calibration is performed on the basis of cross-sectional weights of households.

Adjustment to external data

The survey data are calibrated on the basis of two types of margin.

On the one hand, socio-demographic margins from the $n-1$ *Emploi* (Labour Force) survey to provide a cross-sectional representation of the French population.

On the other hand, total income margins concerning the main sources of income of the French and income vingtile margins taken from the ERFS survey.

At INSEE, all of the household surveys collected during calendar year N are calibrated on the basis of the most recent *Emploi* survey available at the start of year N , or using the margins established on the basis of the 4 quarterly *Emploi* surveys conducted in year $N-1$.

Calibration took place in R using the Icarus package.

The input weights are those established by means of weight sharing. We used the "logit" distance function.

Calibration is done separately for metropolitan France and overseas departments (DOM).

The variables and modalities introduced during calibration are as follows:

- Age cross-referenced with gender in 12 modalities (6 x 2)
- Educational level in 4 modalities- Residential area in 3 modalities (for metropolitan France)
- Socio-professional category in 7 modalities (Indicator for being active in 2 modalities for DOM)
- French nationality indicator in 2 modalities
- Indicator for living in a Priority Neighbourhood in 2 modalities
- Type of household in 5 categories (Indicator for living in a single-person household in 2 modalities for the DOM)
- Total financial income (financial income and property income are aggregated in DOM)
- Total property income (financial income and property income are aggregated in DOM)
- Total wages (wages, unemployment benefits and self-employed income are aggregated in DOM)
- Total pensions
- Total self-employed income (wages, unemployment benefits and self-employed income are aggregated in DOM)
- Total unemployment benefits (wages, unemployment benefits and self-employed income are aggregated in DOM)
- Total income taxes
- The income vingtiles available on ERFS

Final cross sectional weights

After calibration at the household level, the calibrated weight is applied to the individual respondents making up that household.

This gives us a set of cross-sectional weights at the household and individual levels. These two sets of weights are corrected for non-response bias and representative of the French population in $n-1$ onwards.