On recent developments of cross-sectional weighting schemes for children in the Swiss Household Panel

Martina Rothenbühler[1 3], Eric Graf[2], and Alina Matei[2]

[1]FORS, Lausanne, Switzerland
[2]University of Neuchâtel, Switzeralnd
[3]Corresponding author: Martina Rothenbühler, e-mail: martina.rothenbuehler@fors.unil.ch

Abstract

The Swiss Household Panel (SHP) is a longitudinal survey with annual repetition. Each household member aged 14 and over is interviewed individually, whereas information coming from proxy questionnaires is available for younger children. However, no weight is provided for them and it is thus not possible to conduct weighted analysis including the youngest age group. The introduction of children's weights represents therefore a direct gain for the social research in Switzerland, as it enables the production of further knowledge of this age group. From a methodological point of view, the weighting of children is interesting, as only few longitudinal panels include children and there is thus some developments that can be made. We present different cross-sectional weighting systems that are based on the conceptual and theoretical approaches of the current weighting procedures of the SHP as well as on the weighting strategies regarding children in other national household panels. Applying the Generalized Weight Share Method, we show that including the cross-sectional weights of only the original sample members instead of considering also cohabitants leads to estimated frequencies that are closer to known population totals.

Key words: Generalized Weight Share Method, longitudinal survey

1 Presentation of the Swiss Household Panel

The Swiss Household Panel (SHP) is a longitudinal survey with annual repetition and its main objective is the analysis of the socio-economic change within households, in particular the dynamics of the living conditions of the population in Switzerland. Currently, the SHP has two samples. The first sample (SHP_I) started in 1999 with 5074 households, whereas the participants of the second sample (SHP_II) were interviewed for the first time in 2004 (2538 households in 2004). A third sample (SHP_III) is drawn in August 2013 (5000 households). The three samples are representative of the Swiss population.

Within the SHP, we distinguish three different types of participants: Original Sample Members (OSM, members from the first wave), children of OSM (born after the first wave) and non-OSM (cohabitants). Every household member aged at least 14 is eligible to be interviewed, whereas information about children aged under 14 are collected through proxy questionnaires. Until recently, no weight was provided for children.

2 Weighting scheme of the Swiss Household Panel

Weights are used to enable inference to the Swiss population using sample estimates. The weights of the SHP are a mixture of design weights (sampling procedure) and adjustment to non-response as well as calibration (for more details on the weighting scheme see Graf, 2009). Within the SHP there are three different weights:

- Cross-sectional household weights: referring to the households in Switzerland in a given year;
- Cross-sectional individual weights: referring to the Swiss population in a given year;
- Longitudinal individual weights: referring to the Swiss population in 1999 (SHP_I) or 2004 (SHP_II).

Cross-sectional weights are available for all individuals living in a household with at least one OSM. Longitudinal weights are available for OSM only.

When constructing the weights of the SHP, we apply four different theoretical approaches:

- Adjustment for non-response is done by segmentation/Chi-squared Automatic Interaction Detector (CHAID) (Kass, 1980).
- The Generalized Weight Share Method (GWSM) enables to allocate a cross-sectional weight to cohabitants (Lavallée 2007).
- The combination of both panels is performed using the method of Merkouris (2001) that allocates a relative importance to each panel according to its size.
- The weights are calibrated to known population totals by generalized regression according to Deville and Särndal (1992).

3 The Generalized Weight Share Method

The inclusion probabilities of new household members (children of OSM born after the first wave and cohabitants) are not known. As an alternative strategy we can use the inclusion probabilities of OSM and allocate parts of these weights within a household to cohabitants. This procedure, the Generalized Weight Share Method (GWSM), allows to incorporate cohabitants into cross-sectional analyses and to account for an important part of the population dynamics. More generally, the GWSM enables to construct a weight for each unit surveyed in the target population U^B. In the context of longitudinal surveys, the target population U^B represents the population after the first wave. For the construction of cross-sectional weights for children, the target population U^B represents the children in the sample.

We assume that there are links between the units j (individuals) of U^A, from which the sample is selected, and the units k of cluster i (current household composition) of U^B. This relationship is identified by an indicator variable $I_{j,ik}$, where $I_{j,ik} = 1$ if there exists a link

and 0 otherwise. In order to apply the GWSM, each cluster of U^B must have at least one link with U^A. For each unit k of cluster i of U^B, we can calculate the initial weight w'_{ik}:

$$w'_{ik} = \sum_{j=1}^{M^A} I_{j,ik} \cdot \frac{t_j}{\pi_j^A}, \tag{1}$$

where $\sum_{j=1}^{M^A} I_{j,ik}$ represents the sum of the links between unit j of population U^A and unit k of cluster i of U^B, π_j^A is the inclusion probability of unit j and $t_j = 1$ if j is selected in the sample s^A and 0 otherwise. The final weight w_i for units in cluster i is:

$$w_i = \frac{\sum_{k=1}^{M_i^B} w'_{ik}}{\sum_{k=1}^{M_i^B} \sum_{j=1}^{M^A} I_{j,ik}}, \tag{2}$$

where $\sum_{k=1}^{M_i^B} w'_{ik}$ represents the sum of the initial weights for cluster i and $\sum_{k=1}^{M_i^B} \sum_{j=1}^{M^A} I_{j,ik}$ represents the number of links between the units j of U^A and all the units k of cluster i of U^B. This weight corresponds to the average of the sampling weights of the population U^A. The final weight w_i is assigned to all units k within cluster i.

4 Five different cross-sectional weights for children

The construction of the children's cross-sectional weights is based on the GWSM. As mentioned in Section 3, the population U^B, which is sampled indirectly, corresponds to the children in the SHP. However, there are several ways to define U^A:

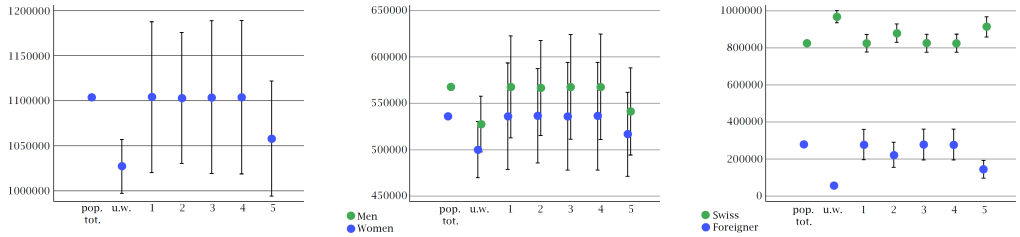
1. *weight_1*: U^A refers to all adults of a household, both OSM and cohabitants, independently of their response status (real and hypothetical weight);
2. *weight_2*: U^A refers to all respondents of a household, both OSM and cohabitants;
3. *weight_3*: U^A is defined by all OSM of a household, independently of their relation to the child;
4. *weight_4*: U^A refers to OSM parents;
5. *weight_5*: U^A corresponds to OSM parents; we allocate half of the weight if there is only one OSM parent.

These five weights are analysed by applying them on different variables. The resulting frequencies are compared with known population totals, but also with unweighted estimates (see Figure 1).

As illustrated in (a) of Figure 1, applying any of the constructed cross-sectional weights 1 to 5 contributes to get population estimations that are closer to the known totals. Analyses without these cross-sectional weights would tend to underestimate the number of children aged under 14 and thus lead to biased estimates if the children are included. The same

is true if we have a look at the estimated totals by gender (b). Using the weights also contributes to get estimations that are closer to the known population totals. The difference between weighted and unweighted frequencies is also visible if the nationality is the variable of interest. In Figure (c) we see that the foreign children are underrepresented in the SHP, whereas the children with a Swiss passport tend to be overrepresented. Again, we can observe that the weighted frequencies are closer to the known population totals than the unweighted estimations.

Figure 1: Difference between weighted and unweighted frequencies



(a) Weighted and unweighted frequencies

(b) Weighted and unweighted frequencies by gender

(c) Weighted and unweighted frequencies by nationality (Swiss/foreigner)

pop. tot.: known population totals u.w.: unweighted totals 1-5: estimated totals with constructed weight

As shown in Figure 1, the known population totals are located in the 95% confidence interval of the weighted frequencies if *weight_1*, *weight_3* or *weight_4* is applied. Because the standard deviations of the frequencies are similar if one of these three weights is applied, the choice regarding which cross-sectional weight should be retained is based on conceptual aspects. The weight share for the cross-sectional weights of the adult cohabitants is performed using the weights of all the OSM of a household, independently of the relation they have with the cohabitants. It seems therefore coherent to use the cross-sectional weights of all of the OSM (*weight_3*) for the construction of the children's weights, as the development of this weight variable is closely related to the current weighting procedure of the SHP.

5 Consequences and future developments

Cross-sectional analyses including the whole sample are now possible within the SHP. However, there are several aspects that need to be improved. First, the inclusion of the children's weights in the datasets slightly affects the allocation factors when combining the panels: the SHP_II is younger and thus gains in relative importance. The allocation factors should be calculated separately for the adults and the children. Second, adjustment to non-response is not yet included in the construction of the weights. Current developments consider the non-response adjustment factor of the reference person (who answered to the proxy questionnaire) on the children's weight. Third, we intend to test the constructed weights using simulation data. Furthermore, the introduction of children's weights seems to be more promising and worthwhile if the number of variables that can be analysed with these weights is larger.

References

Deville, J.-C. and P. Lavallé (2006). Indirect sampling: The foundations of the generalized weight share method. *Survey Methodology 32*(2), 165–176.

Deville, J.-C. and C.-E. Särndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association 87*(418), 376–382.

Graf, E. (2009). Weighting of the Swiss Household Panel. SHP_I wave 9, SHP_II wave 4, SHP_I and SHP_II combined. Detailed description. Technical report, Federal Statistical Office, Neuchâtel, Switzerland.

Kalton, G. and M. J. Brick (1995). Weighting schemes for household panel surveys. *Survey Methodology 21*(1), 33–44.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 29*(2), 119–127.

Latouche, M., J. Dufour, and T. Merkouris (2000). Pondération transversale: combinaison de deux panels ou plus. Technical Report 75F0002MIF00006, Statistics Canada.

Lavallée, P. (2007). *Indirect Sampling*. Springer.

Merkouris, T. (2001). Cross-sectional estimation in multiple-panel household surveys. *Survey Methodology 27*(2), 171–181.