Crossing the age borders in the Swiss Household Panel

Martina Rothenbühler

Swiss Centre of Expertise in the Social Sciences FORS, Lausanne
Presentation of the Swiss Household Panel (SHP)

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 ...

SHP_I

SHP_II

Pilot SHP_III
The Swiss Household Panel (SHP)...

- ... is part of the Swiss Centre of Expertise in the Social Sciences (FORS), located in Lausanne
- ... is founded by the Swiss National Science Foundation
- ... is part of the Cross-National Equivalent File (CNEF)
- ... is a longitudinal survey with annual repetition
- ... has its next data release in October 2012 - the data are freely available
Brief overview of the data

- 2011: 7584 valid individual interviews and 4495 completed household questionnaires (CATI)
- There are between 1500 and 2500 children per wave
  - Information about children comes from proxy questionnaires
  - No weight is allocated to children aged under 14
- Three types of weights
  - Individual cross-sectional weights
  - Individual longitudinal weights
  - Household cross-sectional weights
- Cross-sectional factors that enable "custom made" longitudinal samples
Children’s weights in the German Socio-Economic Panel (SOEP)

- Allocation of cross-sectional weights to cohabitants is performed using the modeling approach:
  - Weights are derived from the household weights.
  - The inclusion probability of an individual corresponds thus to the inclusion probability of the household.
- Children are treated like new entrants and thus receive the household weight.
- The SOEP has a large number of age-specific proxy questionnaires.
Children’s weights in the British Household Panel Study (BHPS)

- There are separate sets of weights for respondent individuals and enumerated individuals.
- Children are treated like cohabitants and the allocation of a cross-sectional weights is thus performed using the Generalized Weight Share Method (GWSM).
  - Average of individual weights of the parents if both parents are original sample members
  - Half of the individual weight if only one parent is an original sample member
- The BHPS has self-completion questionnaires for children aged 10 to 15.
Overview of the different approaches of the weighting procedure in the SHP

- 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 according Deville and Särndal (1992). They suggest to apply a calibration by generalised regression.
Within the SHP, we use the GWSM to allocate a cross-sectional weight to non-original sample members, of whom we don’t know the inclusion probability.

The GWSM can also be used to construct children’s weights.

Basically, the GWSM produces an estimation weight for each unit surveyed in the target population $U^B$ (children). This estimation weight corresponds to the average of the sampling weights of the population $U^A$ from which the sample is selected.
The Generalized Weight Share Method (GWSM)

We start with a sample $s^A$ containing $M^A$ units, where $\pi^A_j$ represents the inclusion probability of unit $j$. The target population $U^B$ contains $M^B$ units and is divided into $N$ clusters (for example households), each of them containing $M^B_i$ units.

In order to apply the GWSM, each cluster (household) of $U^B$ must have at least one link with $U^A$, that is:

$$L^B_i = \sum_{k=1}^{M^B_i} \sum_{j=1}^{M^A} l_{j,ik} > 0$$
For each unit $k$ of cluster $i$ of $\Omega^B$, we can calculate the initial weight $w'_{ik}$ as follows:

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

where $t_j = 1$ if $j \in s^A$ and 0 otherwise.
For each unit $k$ of cluster $i$ of $\Omega^B$, we get the total number of links $L_{ik}^B$:

$$L_{ik}^B = \sum_{j=1}^{M^A} l_{j,ik}$$

$L_{ik}^B$ represents the number of links between the units $j$ of $U^A$ and the unit $k$ of cluster $i$ of $U^B$. The quantity $L_i^B = \sum_{k=1}^{M^B_i} L_{ik}^B$ corresponds to the number of links present in cluster $i$. 
The Generalized Weight Share Method (GWSM) - Step 3

Now, we can calculate the final weight $w_i$:

$$w_i = \frac{\sum_{k=1}^{M_i^B} w_{ik}'}{\sum_{k=1}^{M_i^B} \sum_{L_{ik}^B}}$$

We calculate the final weight of each cluster $i$, which is obtained by calculating the ratio of the sum of the initial weights for the cluster and the total number of links for that cluster.
Finally, we assign $w_{ik} = w_i$ for all $k \in U^B$. The final weight is assigned to all units $k$ within the cluster $i$. 

Martina Rothenbühler
Crossing the age borders in the Swiss Household Panel
Five types of children’s cross-sectional weights

- The population $U^B$, which is sampled indirectly, corresponds to the children in the SHP in all the constructed weights.
- However, the population $U^A$ differs:
  1. $U^A$ corresponds to all the adults within a household, both original sample members and cohabitants, independently of the response status (real and hypothetical weight)
  2. $U^A$ corresponds to all the respondents within a household, both original sample members and cohabitants
  3. $U^A$ corresponds to all the original sample members within a household, independently of their relation to the child
  4. $U^A$ corresponds to the OSM parents
  5. $U^A$ corresponds to the OSM parents; we allocate half of the weight if there is only one parent present (equal to the BHPS)
Difference between weighted and unweighted frequencies

0: known population totals
1-5: estimated total with different weight types
6: unweighted totals
Difference between weighted and unweighted frequencies by gender

0: known population totals
1-5: estimated total with different weight types
6: unweighted totals

Martina Rothenbühler
Crossing the age borders in the Swiss Household Panel
Difference between weighted and unweighted frequencies by nationality

0: known population totals
1-5: estimated total with different weight types
6: unweighted totals
Comparison between known population totals and the weighted frequencies using weight type 1

<table>
<thead>
<tr>
<th>Var.</th>
<th>N</th>
<th>n</th>
<th>Weight. freq.</th>
<th>Std. dev.</th>
<th>95% conf. interval</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1097820</td>
<td>1528</td>
<td>1096418</td>
<td>37361</td>
<td>[1023093 ; 1169743]</td>
<td>1402</td>
</tr>
<tr>
<td>Men</td>
<td>564592</td>
<td>802</td>
<td>563705</td>
<td>26151</td>
<td>[512380 ; 615030]</td>
<td>887</td>
</tr>
<tr>
<td>Women</td>
<td>533229</td>
<td>726</td>
<td>532713</td>
<td>25965</td>
<td>[481754 ; 583672]</td>
<td>516</td>
</tr>
<tr>
<td>Swiss</td>
<td>824653</td>
<td>1443</td>
<td>836177</td>
<td>23660</td>
<td>[789742 ; 882611]</td>
<td>11524</td>
</tr>
<tr>
<td>Foreign</td>
<td>273167</td>
<td>85</td>
<td>260241</td>
<td>36264</td>
<td>[189069 ; 331414]</td>
<td>12926</td>
</tr>
</tbody>
</table>
Comparison between known population totals and the weighted frequencies using weight type 2

<table>
<thead>
<tr>
<th>Var.</th>
<th>N</th>
<th>n</th>
<th>Weight. freq.</th>
<th>Std. dev.</th>
<th>95% conf. interval</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1097820</td>
<td>1499</td>
<td>1103296</td>
<td>35652</td>
<td>[1033323 ; 1173268]</td>
<td>5476</td>
</tr>
<tr>
<td>Men</td>
<td>564592</td>
<td>788</td>
<td>566961</td>
<td>25517</td>
<td>[516881 ; 617042]</td>
<td>2369</td>
</tr>
<tr>
<td>Women</td>
<td>533229</td>
<td>711</td>
<td>536334</td>
<td>24939</td>
<td>[487388 ; 585280]</td>
<td>3105</td>
</tr>
<tr>
<td>Swiss</td>
<td>824653</td>
<td>1417</td>
<td>870749</td>
<td>24471</td>
<td>[822721 ; 918778]</td>
<td>46096</td>
</tr>
<tr>
<td>Foreign</td>
<td>273167</td>
<td>82</td>
<td>232546</td>
<td>33504</td>
<td>[166789 ; 298304]</td>
<td>40621</td>
</tr>
</tbody>
</table>

Martina Rothenbühler

Crossing the age borders in the Swiss Household Panel
### Comparison between known population totals and the weighted frequencies using weight type 3

<table>
<thead>
<tr>
<th>Var.</th>
<th>N</th>
<th>n</th>
<th>Weight. freq.</th>
<th>Std. dev.</th>
<th>95% conf. interval</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1097820</td>
<td>1484</td>
<td>1097820</td>
<td>39681</td>
<td>[1019940 ; 1175701]</td>
<td>0</td>
</tr>
<tr>
<td>Men</td>
<td>564592</td>
<td>780</td>
<td>564592</td>
<td>27239</td>
<td>[511130 ; 618054]</td>
<td>0</td>
</tr>
<tr>
<td>Women</td>
<td>533229</td>
<td>704</td>
<td>533229</td>
<td>27241</td>
<td>[479764 ; 586693]</td>
<td>0</td>
</tr>
<tr>
<td>Swiss</td>
<td>824653</td>
<td>1402</td>
<td>824654</td>
<td>23002</td>
<td>[779508 ; 869799]</td>
<td>1</td>
</tr>
<tr>
<td>Foreign</td>
<td>273167</td>
<td>82</td>
<td>273167</td>
<td>39385</td>
<td>[195866 ; 350467]</td>
<td>0</td>
</tr>
</tbody>
</table>
Comparison between known population totals and the weighted frequencies using weight type 4

<table>
<thead>
<tr>
<th>Var.</th>
<th>N</th>
<th>n</th>
<th>Weight. freq.</th>
<th>Std. dev.</th>
<th>95% conf. interval</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1097820</td>
<td>1466</td>
<td>1097820</td>
<td>40674</td>
<td>[1017988 ; 1177653]</td>
<td>0</td>
</tr>
<tr>
<td>Men</td>
<td>564592</td>
<td>767</td>
<td>564592</td>
<td>27676</td>
<td>[510272 ; 618912]</td>
<td>0</td>
</tr>
<tr>
<td>Women</td>
<td>533229</td>
<td>699</td>
<td>533229</td>
<td>27733</td>
<td>[478795 ; 587662]</td>
<td>0</td>
</tr>
<tr>
<td>Swiss</td>
<td>824653</td>
<td>1385</td>
<td>824653</td>
<td>24428</td>
<td>[776708 ; 872598]</td>
<td>0</td>
</tr>
<tr>
<td>Foreign</td>
<td>273167</td>
<td>81</td>
<td>273167</td>
<td>39629</td>
<td>[195386 ; 350948]</td>
<td>0</td>
</tr>
</tbody>
</table>
Comparison between known population totals and the weighted frequencies using weight type 5

<table>
<thead>
<tr>
<th>Var.</th>
<th>N</th>
<th>n</th>
<th>Weight. freq.</th>
<th>Std. dev.</th>
<th>95% conf. interval</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1097820</td>
<td>1466</td>
<td>1097820</td>
<td>41765</td>
<td>[1015847 ; 1179793]</td>
<td>0</td>
</tr>
<tr>
<td>Men</td>
<td>564592</td>
<td>767</td>
<td>564592</td>
<td>27999</td>
<td>[509638 ; 619545]</td>
<td>0</td>
</tr>
<tr>
<td>Women</td>
<td>533229</td>
<td>699</td>
<td>533228</td>
<td>28464</td>
<td>[477362 ; 589095]</td>
<td>1</td>
</tr>
<tr>
<td>Swiss</td>
<td>824653</td>
<td>1385</td>
<td>824653</td>
<td>26128</td>
<td>[773372 ; 875934]</td>
<td>0</td>
</tr>
<tr>
<td>Foreign</td>
<td>273167</td>
<td>81</td>
<td>273167</td>
<td>39653</td>
<td>[195340 ; 350994]</td>
<td>0</td>
</tr>
</tbody>
</table>
Comparison of the five types of weight

- Including all the adults, independently of the response status or the presence at the time the sample was selected (type 1 and 2), leads to small standard deviations, but to (more) important differences between known and estimated population totals.

- Including only original sample members (types 3 to 5) leads to slightly larger standard deviations, but (almost) equal population totals.

- In order to stay in line with the current weighting procedure, we decided to follow type 3 of the cross-sectional children’s weights, where all the original sample members within a household are considered.
Conclusion: Consequences of weighting children

- The inclusion of children’s weights in the datasets affects slightly the allocation factors when combining the panels: the second panel is younger and thus gains in relative importance.

- The introduction of enumeration weights thus seems to be promising.

- The construction of longitudinal weights for children is worth to try, but affords a change of definition.

- The introduction of children's weights seems to be more promising and worthwhile if the number of variables that can be analyzed with these weights is larger.

- (Cross-sectional) Analyses including the whole sample are now possible.
References

Indirect sampling: The foundations of the generalized weight share method.
*Survey Methodology* 32(2), 165–176.

Calibration estimators in survey sampling.

Weighting schemes for household panel surveys.
*Survey Methodology* 21(1), 33–44.

An exploratory technique for investigating large quantities of categorical data.

Pondération transversale: combinaison de deux panels ou plus.

*Indirect Sampling.*
Springer.

Cross-sectional estimation in multiple-panel household surveys.
*Survey Methodology* 27(2), 171–181.
Thank you for your attention!

www.swisspanel.ch
www.fors.unil.ch