Constructing Cross-Sectional Weights for the German Panel on Household Finances

Panagiota Tzamourani[1]

[1]Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, DE-60431, Germany

Abstract

The paper describes the construction of cross-sectional weights for the second wave of the Panel on Household Finances survey, the German part of the euro area Household Finance and Consumption Survey. Its first wave was conducted in 2010/2011 and its second in 2014, the last including both the respondents of the previous wave and a refresher sample. The construction of cross-sectional weights in a panel survey poses certain challenges as changes in the participating households, changes in the population and panel attrition need to be taken into account. To adjust the weights of the panel households we use a 'base weight' approach as described for example in Verma, Betti and Ghellini (2006). The proposed method relies on the construction of person weights, from which household weights are derived. Non-response adjustments follow, which are carried out separately for the panel, split and refresher samples. At the final stage the three components are merged and their weights are calibrated together.

Key Words: panel survey, cross-sectional weights, PHF, HFCS

1. Introduction

The paper describes the construction of cross-sectional weights for the second wave of the Panel of Household Finance (PHF) survey. The PHF is a representative survey of German households encompassing questions on households' income, assets, debt, saving and consumption. It is conducted by the Bundesbank and it constitutes the German part the Household Finance and Consumption Survey (HFCS) conducted in the euro area. It first took place in 2010-11 and was repeated in 2014, comprising a panel and a refresher component. The survey, as most panel surveys, has the aim to provide both estimates of change over time and cross-sectional estimates, relevant to the years the survey took place. Using a panel sample to provide cross-sectional estimates poses certain difficulties. A panel sample is subject to changes over time: households change their composition as sample members move in and out of the household; sample members drop out of subsequent waves of the survey; moreover, the population changes over time and a panel sample is not considered representative of the population after two or three years from the time it was drawn. Weighting, and the integration of the refresher component if there is one available, aims to take into account these changes over time and make the (joint) sample representative of the population at each point in time the survey was carried out.

The paper will describe the construction of cross-sectional weights for second wave of the PHF, with emphasis on the adjustments of the weights of the panel component. Revisions are possible until the release of the data, envisaged to take place in mid-2016.

The rest of the paper is structured as follows: Section 2 will describe the PHF survey in more detail. Section 3 will present the general framework for the construction of cross-sectional weights for a panel survey. Section 4 will describe the actual steps of constructing the cross sectional weights for the PHF. Section 5 will conclude.

2. The Panel on Household Finances (PHF)

The PHF is a representative survey of German private households. The questionnaire comprises questions on households' real and financial assets, inheritances, debts, income, saving, consumption and furthermore for all the household members above 16 years old questions about their employment, income and pension entitlements. Questions on expectations, risk attitudes and financial literacy are also included. An overview of the first wave of the survey is given in von Kalkreuth et al (2013). The survey is conducted by means of a computer-assisted personal interview (CAPI). The fieldwork in the first two waves has been conducted by the research institute infas[1].

Sampling is based on a multistage stratified design. The sampling frame is the municipality register ('Einwohnermelderegister') and oversampling of the wealthy is employed. Because of this particular feature, the stratification is based on income and wealth characteristics of geographical areas. Three strata are formed: wealthy rural municipalities, other rural municipalities and cities. At the first stage municipalities from the first two strata and cities from the third stratum are drawn. At a second stage, for the selected cities only, street sections are stratified in wealthy and other, and street sections are drawn. In the final stage persons from the municipality register are drawn in the selected municipalities or street sections and the households the selected persons belong to are considered the final sampling units. Households are oversampled in wealthy small municipalities and in wealthy street sections in cities.

In the first wave of the survey (2010/2011) 20,501 households were initially selected in the sample from which 3,565 took part in the survey. In the second wave, in 2014, the gross sample constituted of the 3,202 panel households that had agreed to be contacted in a subsequent wave of the survey, and 12,805 refresher households. The sampling design upon which the selection of the refresher households is based is similar to the one of the first wave[2]. The net sample comprised 2,151 panel households and 2,270 refresher households.

As panel households change over time it is essential to establish 'follow' rules, to determine which household members will be followed in subsequent waves. The PHF is similar to the rules employed by the 'Panel Study on Income Dynamics (PSID)' (Heeringa et al., 2008). According to these rules, all households with at least one 'original sample member' are to be followed, that is, form part of the gross sample of the subsequent waves (until they explicitly request to drop out of the survey, or move

---

[1] infas Institut für angewandte Sozialwissenschaft GmbH, Bonn, Germany.
[2] The survey design is described in detail in infas (2013) whereas Schmidt and Eisele (2013) provide an evaluation of the oversampling strategy.

permanently out of the 'in-scope' population). Original sample members are all the members of the wave 1 households, plus the newborn children and the adopted children (fulfilling certain conditions). Non-original sample members, as for example, people moving in with an original sample member, are only followed as long as they belong to the household of the original sample member. Original sample members that move out of the wave 1 household they belonged to, form the 'split' households and are followed as the other original sample members.

3. Cross-sectional weights in panel setting

The aim of cross-sectional weighting in a panel setting is to allow inference about the total 'in-scope' population at a given date using all samples. The in-scope population for the PHF, as for most household surveys, is the total of private households in the country, i.e. persons living in communal housing are excluded. As the in-scope population is enriched through births, immigration, people leaving communal housing and moving into regular housing, part of the population is not captured by the samples chosen in previous years. The refresher sample in the PHF aims to, besides increasing the overall sample size, to account for these changes. Moreover, the inclusion in the sample of all household members that moved-in with panel members since the previous realization of the survey also helps to take into account of the changing population. Similarly, the sample and the 'in-scope' population are affected as people die, move out of the country or into communal housing.

While though the panel members living the population do not present particular difficulties for the weighting (Ardilly and Lavallee, 2007) entrants in the population may pose a need for adjusting selection probabilities and weights. Different methods have been proposed to make these adjustments and different panel surveys may employ different methodologies. Methods differ as to whether they are applied to households as a whole or whether they are applied to person weights from which household weights are computed. They may be based on estimating selection probabilities with regression models (as it is done, for example, for the HILDA survey (Watson, 2012) or based on weight sharing methods, as for example described in Lavalle, P. (1995), Merkouris (1999), Rendtel and Harms (2009). Schonlau, Kroh and Watson (2013) compare the various methods and how they are used in various surveys. Our methodology stems from the 'base weight' approach, as described in Verma, Betti and Ghellini (2006), which is essentially a weight-share method.

In the following we will present all the steps leading to the construction of cross-sectional weights for the PHF, part of which also entails the adjustments to the weights of the panel households.

4. Cross-sectional weights for the PHF

The construction of the weights entails the following steps: computation of design weights for the refresher households; potential adjustments to the weights of the panel households to account for changes in the household composition-and the differential selection probability they imply; non-response adjustments for the samples of the panel,

split and refresher households; merging the samples together; calibration of the merged household and person weights. In the following we will present each of these steps[3].

4.1 Design weights

The design weights are based on the sampling design and aim to correct for the differential selection probabilities of households. The design weights are defined as the inverse of the selection probability of each sample unit. In a multistage design the selection probability of the final sample unit is the product of the relevant selection probabilities at each stage.

The design weights for the PHF are computed exclusively at infas and a more detailed description of the procedure is published in their methodological report (infas 2015).

4.2 Adjustments for the panel households

Panel households carry the weight computed from the first wave of the survey. This can be considered the 'design' weight. However, as discussed above, some adjustments for changes in the composition of the participation households need to be made.

4.2.1 The 'base'-weight approach

Following Verma, Betti and Ghellini (2006) we proceed as follows: The initial, 'base', cross-sectional weight of a household member is the final cross-sectional weight of the household it belonged to in wave 1. Adjustments to these weights are made, based on whether any entrants to the household could have been selected in the sample at the time of wave 1. This would have been possible if they then lived in Germany, in a private household, and if they were at least 18 years old at the time. We approximate this definition, as we do not have information on whether they lived in a private household or not, using the variable indicating whether the respondent lived in Germany in 2011, the year where most of the interviews of the first wave took place and the respondent's age. In particular, the following adjustments are made:

Newborns take the weight of the mother, which is equal to the household weight of wave 1 she belongs to.

For persons moving in sample from outside the population of wave 1, the new persons take the household weight of the household they move in to.

For the persons moving in the sample from within the population of wave 1 (if they were in Germany in 2011 and were at least 18 years old at the time), their base weight is set to 0. In that case, assuming there are no other changes in the household, this new household weight will be reduced compared to weight in wave 1. In our data there were 131 panel households falling under this case.

Finally, for persons that died or moved out of the population of wave 2 their base weight is set to missing so that they are disregarded from further adjustments.

Original sample members that moved within the population are followed (split-households). They carry their final weight of wave 1.

At the end, household weights are computed as the average of the adjusted person weights in the household. The undertaken changes and their effect on the household weight is summarised in Table 1[4].

---

[3] The construction of the weights is largely undertaken by infas (see infas(2013) and infas(2015)) though modifications at all the adjustments after the construction of the design weights are undertaken by the Bundesbank.

Table 1: Adjustments to person -'base'- weights for changes in the household composition and their effect on the household and person weights

Changes in the household	Intermediate person weight	Household weight	Person weights
Household entrants part of population in Wave 1	receive 0 weight	downweighted	downweighted
Household entrants not part of population in Wave 1 ('immigrants')	receive household weight	remains same	remain same
Births/adoptions	receive household weight	remains same	remain same
Household split	members carry household weight	remains same	remain same
Merging households	as household entrants part of population in wave 1	downweighted	downweighted
Death	set to missing	remains same	remain same

4.2.2 An alternative approach

An alternative approach to account for entrants in the household that were part of the population in wave 1, is to correct for their increased underlying probability of selection by adding to their wave 1 selection probability the overall probability of being selected in the sample (n/N). This is being used for the PASS survey in Germany (Trappmann, M. (2013) and has also been proposed by the infas institute for the PHF[5]. In that case, the adjusted household weight (wh_{w2}) for panel households that have entrants from within the in-scope population of wave 1 is computed as follows:

$wh_{w2} = 1 / ((1/wh_{w1}) + (n_{w1}/N_{w1}) - (1/wh_{w1}) * (n_{w1}/N_{w1}))$,

where wh_{w1} is the final household weight from wave 1, n_{w1} is the sample size from wave 1 (3565 households) and N_{w1} is the corresponding population (39.673.000 households). The second part of the denominator is a small correction factor to allow for the probability of double selection.

As in the 'base'-weight approach, the household weight for the households that this adjustment applies is decreased.

The two alternative methodologies for the adjustments of the panel households described above typically result in adjustments of different size for the households affected. In our data the adjustments were typically larger with the 'base'weight method for the weights of one person households (as of wave 1) with a new entrant in wave 2 (and new in the population of wave 2) and smaller in the rest of the cases. The calibration we employed,

---

[4] Schonlau, Kroh and Watson (2013) provide similar tables for the BHPS, SHP, PSID, HILDA and SOEP.
[5] Infas (2015).

which is described further below, has on the whole contracted the discrepancies between the calibrated weights coming from the two different methods. The effect on the estimates of the main substantive variables is currently being examined.

4.3 Non-response adjustments

Non-response adjustments aim to improve the representativeness of the sample and account for some of the selectivity of non-response. One of the methods to compute such adjustments is to build logit propensity models. For our purposes separate models are defined for the panel, split and refreshment households. The variables used are considered to be correlated with the non-response mechanism and must be available for both respondents and non-respondents. The variables fitting these criteria available in our data are the following:

- gender, age class and nationality of the person selected from the municipality register
- paradata collected by the interviewer upon their first visit to the selected household's dwelling. In particular, the following variables are used: the type of building the households resides, the interviewer's rating of the dwelling, the interviewer's rating of the dwelling in comparison to the neighborhood, rating of the neighborhood, existence of security measures
- Bundesland and municipality size class

In addition, for the panel and split households household characteristics from the first wave are also available. From those we used the household size, homeownership status, age, gender, education and work status of the reference person (or the person moving out in the case of split households).

After fitting these models[6], the existing household weights, that is the design weights for the refresher sample and the adjusted weights from the procedure described in Section 4.2 are multiplied with the inverse of the predicted response propensity from the fitted models.

4.4 Merging of the sample components

In order to proceed with the calibration of a unified sample, the panel (including the split households) and the refresher samples are merged and their weights are multiplied with the proportion of the corresponding sample component (panel/refresher) in the total sample. After the merging, in order to avoid excess variation, weights are trimmed, setting the values of the weights below and beyond the 5th and 95th percentile of the distribution to the respective percentile values.

4.5 Calibration

The purpose of the calibration is to adjust the weights so that aggregates computed from the sample match known population totals. Both household and person weights are calibrated. First household weights are calibrated. The calibrated weights are distributed to the household members and person weights are also calibrated so that the aggregates relating to the persons also match known population totals. We further iterate this procedure a few times, that is we average the calibrated person weights and repeat the household calibration. At the final step we calibrate households and distribute the weights to the persons without further calibrating. In that way we have person weights equal the

---

[6] The basic structure of the non-response models has been proposed by infas (infas(2015), to appear).

household weights while both household and person totals approximate well their population targets. The population totals are taken from the Microcensus[7].

For the calibration of households the following variables are used:

- Combination of age and gender of the main income earner
- Combination of age and education level of the main income earner
- Work status of the main income earner
- Nationality (German / non-German)
- Household size
- Homeownership status
- Size of dwelling
- Homeownership status combined with dwelling size
- Number of households per Bundesland
- Number of households per municipality size class

For the calibration of persons the following variables are used:

- Age combined with gender
- Age combined with education level
- Age for East and West Germany
- Work status
- Nationality (German/ non-German)
- Population per Bundesland
- Population per municipality size class
- Population per municipality size class for East and West Germany

5. Conclusions

The construction of cross-sectional weights for panel surveys poses particular methodological difficulties and assumptions. In this paper we described the methodology employed to construct cross-sectional household and person weights for the Panel of Household Finances, laying out the treatment of panel households, the non-response adjustments and calibration employed. The effect of different methods of adjusting for the changes in the panel households is the subject of ongoing research.

Acknowledgements

---

[7] The microcensus is a sample survey covering roughly 1% of the population in Germany each year. It is carried out by DESTATIS, the German Federal Statistical Office.

References

Ardilly P. And P. Lavallee (2007). Weighting in rotating samples: The SILC survey in France. Survey Methodology, Vol. 33 pp. 131-137.

Heeringa E.S., K. McGonagle, R. Schoeni, F. Stafford (2008) Panel study of income dynamics revised longitudinal weights 1993—2005 (p.6-8), Frank, Panel study of income dynamics technical paper series, University of Michigan, Ann Arbor, USA.

Infas (2013) Methodenbericht zur zweiten Erhebungswelle der Studie Private Haushalte und ihre Finanzen, Bonn. Available at http://www.bundesbank.de/Navigation/EN/Bundesbank/Research/Panel_on_household_finances/Methodology/methodology.html

Infas (2015) Methodenbericht zur zweiten Erhebungswelle der Studie Private Haushalte und ihre Finanzen, Bonn. To appear online.

Lavalle, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. Survey Methodology, 21,pp. 25-32.

Merkouris., T. (1999). On the Weight Share Method for Panel Household Surveys. Proceedings of the Survey Research Methods Section, ASA

Trappmann, M. (2013). Weighting. In Bethmann. A., B. Fuchs, A.Wurdack (2013), User Guide, Panel Study Labour Market and Social Security (PASS), Wave 6.

Rendtel U. and T. Harms (2009). Weighting and calibration of household panels. In P.Lynn (Ed.), Methodology of Longitudinal Surveys. Chichester: John Wiley and Sons.

Schmidt, T. and M. Eisele (2013), Oversampling vermögender Haushalte im Rahmen der Studie "Private Haushalte und ihre Finanzen (PHF)", in: T. Riede, S. Bechtold and N. Ott (edt.), Weiterentwicklung der amtlichen Haushaltsstatistiken, 31-40, Scivero Verlag, Berlin.

Schonlau M., M. Kroh and N. Watson (2013). The implementation of cross-sectional weights in household panel surveys. Statistics Surveys. Vol. 7, pp 37-57.

Verma, V. G. Betti, G. Ghellini (2006). Cross-sectional and longitudinal weighting a rotational household panel: Applications to EU-SILC. Working Paper n. 67, December 2006

von Kalckreuth, U., M. Eisele, J. Le Blanc, T. Schmidt and J. Zhu (2012), The PHF: A comprehensive panel survey on household finances and wealth in Germany. Bundesbank Discussion Paper No. 13, Deutsche Bundesbank, Frankfurt am Main.

Watson, N. (2012). Longitudinal and Cross-sectional weighting. Methodology for the HILDA Survey. HILDA PROJECT TECHNICAL PAPER SERIES. No. 2/12, December 2012