Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers
on Multidisciplinary Panel Data Research

313

Matthias Schonlau • Martin Kroh

On the Equivalence of Common Approaches to Cross Sectional Weights in Household Panel Surveys

Berlin, August 2010

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

On the Equivalence of Common Approaches to Cross Sectional Weights in Household Panel Surveys

Matthias Schonlau[1,2] and Martin Kroh[1,3]

[1] SOEP (German Socio Economic Panel Study) at DIW Berlin (German Institute for Economic Research), Germany

[2]RAND Corporation, Pittsburgh, USA

[3] University of Bamberg

Corresponding author:

Matthias Schonlau, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA, 15213, USA

matt@rand.org

Abstract

The computation of cross sectional weights in household panels is challenging because household compositions change over time. Sampling probabilities of new household entrants are generally not known and assigning them zero weight is not satisfying. Two common approaches to cross sectional weighting address this issue: (1) "shared weights" and (2) modeling or estimating unobserved sampling probabilities based on person-level characteristics. We survey how several well-known national household panels address cross sectional weights for different groups of respondents (including immigrants and births) and in different situations (including household mergers and splits). We show that for certain estimated sampling probabilities the modeling approach gives the same weights as fair shares, the most common of the shared weights approaches. Rather than abandoning the shared weights approach when orphan respondents (respondents in households without sampling weights) exist, we propose a hybrid approach; estimating sampling weights of newly orphan respondents only.

Key words: BHPS, HILDA, PSID, SOEP, modeled weights, shared weights, fair shares

Acknowledgements

This work was done while the first author, M. Schonlau, was on sabbatical with the SOEP group at the DIW Berlin. Primary funding for this work came from the SOEP group and for the sabbatical as a whole in addition to a fellowship of the Max Planck Institute for Human Development (MPIB, Berlin). We thank Gert G. Wagner for encouraging this work.

1. Introduction

Household panel surveys are sample surveys in which the same private households are interviewed repeatedly over time (e.g. once a year). They are typically general purpose surveys with multiple topics, and have become an important source of socio economic and other micro data. Many countries around the world are financing household survey panels, including USA (PSID, http://psidonline.isr.umich.edu), Great Britain (BHPS, http://www.iser.essex.ac.uk/survey/bhps), Germany (SOEP, http://www.diw.de/en/soep), Canada (SLID, http://www.statcan.gc.ca/), Australia (HILDA, http://www.melbourneinstitute.com/hilda), Switzerland (SHP, http://www.swisspanel.ch), Netherlands (LISS, http://www.centerdata.nl/en/TopMenu/Projecten/MESS), and Chile (CASEN, http://mideplan.cl/casen). South Africa's household panel (http://www.nids.uct.ac.za) is in wave 2, the Israel Central Bureau of Statistics is about to set up a household survey panel for Israel (Thomas Caplan, Israel Central Bureau of Statistics, personal communication).

Household panel surveys differ from one-time cross sectional surveys. Time affects household panel surveys in two ways: (a) the target population changes over time, (b) the household composition may change over time. In all high quality panels the sample in wave 1 is a probability sample of a target population (e.g. German private households) at one point in time. Because of immigration, emigration, births and deaths the target population in the following year is slightly different and those differences accumulate over time. It is possible to take the view that the purpose of the panel is to follow the population in year 1 over time, thereby eliminating the need to address immigration and births. However, this view is not satisfactory in practice as it does not allow for cross sectional analyses except for the first wave. Compounding this issue, the household composition changes over time as a consequence of marriages/cohabitation, separations/divorces, births, adoptions, deaths, children moving out. This raises the question of how to compute weights for new household entrants, i.e. respondents who move into existing households.

The sampling probability of new household entrants is usually unknown. A related issue is the effect of individual persons moving out of a household on weights. Depending on specific so-called following rules, some respondents are traced as they form their own new households whereas others are not traced. The most common situations in which new households are formed are (1) the separation of the head of household from his partner, and (2) grown children moving out. From a substantive point of view, following "movers-out" is desirable because in this case a more complete story about population dynamics can be told with the panel data.

We compare the implementation of cross sectional weights of several household survey panels and derive conditions under which the two most common approaches are equivalent; i.e. lead to the same cross sectional weights. The next section introduces the two most common approaches to cross sectional weighting, "shared weights" and "modeling". Section 3 contains a comparison of how these approaches are implemented in several large household surveys. Section 4 gives conditions under which the weights of one of the "shared weights" approaches coincide with the "modeling approach". Section 5 addresses the issue that the shared weights approach does not work if none of the household members has a sampling weight and proposes one possible solution. Section 6 concludes with a discussion.

2. Cross-sectional weights for new household entrants

When new household members enter a household panel after wave 1, it is common to compute their cross-sectional weights. The other option, assigning no weight, is not desirable because it wastes data in the sense that respondents deliver information which gets a weight of zero.

The cross-sectional weight is computed from the probability of selection in wave 1. The probability of selection for new household entrants depends on their household membership history over the life of the panel (Lynn 2009, p.28). However, the membership history of new household entrants prior to their entry is often unknown. For example, suppose in the second wave of a panel survey person A moves into a household that consists only of person B. Then there are two paths though which a household may be included in wave 2: by sampling person A or person B in wave 1 (or both). To properly compute the household weight for wave 2, one needs to compute the probability of sampling A or B in wave 1. Failure to make a correction would overstate the number of households with new entrants (Watson 2004) .

Difficulties arise because the wave 1 selection probabilities for new household entrants are typically not known. One approach is to estimate these probabilities which we call the modeling approach (Galler 1987, p.313). A completely different solution is the “shared weights” approach (Ernst 1989) which includes the “fair shares” approach . The PSID considers only members of wave 1 and their children to be sample members and implicitly assigns weight 0 to other cohabitants (non-sample members).

The "Shared weights" approach

"Shared weights" (Ernst 1989; Kalton and Brick 1995; Lavallée 1995; Lavallée and Caron 2001; Rendtel and Harms 2009) is a strategy for developing cross sectional weights that only requires selection probabilities of individuals selected in the original sample. The "shared weights" approach keeps the sum of individual weights within a household constant, redistributing the weights among the individuals as new individuals enter a household. For example, suppose each person in a two-person household has weight 1500. When a new entrant joins this household, the total weight of 3000 is redistributed among 3 people giving each person a weight of 1000. When the total weight is distributed evenly among the household members, this is called the "equal person weight" or "fair shares" method to redistributing weights. Other weight sharing schemes exist (Rendtel and Harms 2009).

It turns out that the redistribution of weights among household members yields unbiased estimators of a population total, though efficiency depends on the sampling probabilities and cannot always be assessed (Kalton and Brick 1995). However, the shared weights approach requires that at least one wave 1 respondent still lives in the household (Kalton and Brick 1995, p. 3-1, 6-1; Lynn 2009, p.28) . This means that associated persons that leave a household – such as a spouse who joined the household in wave 2 and who later divorced and moved out - receive zero weight. This is unproblematic when only wave 1 sample members and their children are followed as is the case in BHPS[1]. This is not acceptable when wider following rules are adopted as is the case in the SOEP, HILDA, and the Swiss Household Panel (SHP) (after wave 9).

---

[1] BHPS also follows parents of children who have at least one wave 1 parent. They are assigned a weight of zero when living in households without a weight to share.

The shared weight approach implies the sum of the individual weights remain constant over time except for additional weight due to new birth/ adoption and recent immigrants. In practice, however, household weights vary from year to year because of corrections for nonresponse and poststratification.

The modeling approach

Even though the earlier history of entrants in later waves is not known, it is possible to model the individual selection probabilities, e. g. via regression. The probability of selecting household H_i is the probability of selecting one or more constituent households:

$$P(H_i) = P(h_1 \cup h_2 \cup h_3 \cup \cdots \cup h_k) = 1-(1-p_1)(1-p_2)\cdots(1-p_k) \qquad (1)$$

where $h_1, \ldots, h_k$ are the constituent households in wave 1 which jointly form the new household at a later wave, and where $p_1,\ldots,p_k$ are the corresponding selection probabilities. Equation (1) assumes independence between the constituent households (Kalton and Brick 1994, Equation 3.3). A constituent household or a group of household entrants refers to entrants that moved together from their old household to the new household (e.g., a mother with children). Overall, the independence assumption appears reasonable even though it might not hold, for example, because people who get married might be geographically clustered. The selection probability of the household that was in the original sample, p_1, is known but the selection probabilities corresponding to new entrant groups are unknown because they were not part of the original sample. This approach is implemented in HILDA (Watson 2004).

The approach taken in the SOEP (Galler 1987) has two modifications. First, equation (1) is simplified by removing joint probabilities. Neglecting the smaller joint probabilities, equation (1) can be rewritten as

$$P(\text{selecting HH}) = p_1 + p_2 + \ldots + p_k \quad (2)$$

Second, SOEP does not consider groups of new entrants (or constituent households) but rather treats each new entrant as a separate unit. For example, if a mother and her grown child move into a respondent household, SOEP would compute the selection probability as $p_1 + p_2 + p_3$ (probabilities corresponding to the original household, the mother and the child) whereas equation (1) implies $p_1 + p_2 - p_{12}$ (probabilities corresponding to the original household, the mother-child household, and the joint probability of selecting both households).

In both approaches, unknown probabilities need to be estimated. This is done via regression analyses. SOEP uses ordinary least squares regression with logit(p) as a dependent variable. The SOEP regressions explain about 90% of the variation (R^2=0.9) for early waves and about 50% of the variation (R^2=0.5) for recent waves. Weights are computed as the inverse selection probabilities (Horvitz and Thompson 1952) which are derived from the regression results. Therefore, as a new group moves into a household, the selection probability of the combined household increases and the weight of the combined household decreases.

3. Implementation of cross sectional weights in survey panels

The two basic approaches to cross sectional household weights outlined in the previous section have been implemented across a variety of household panel surveys. We consider the effect of new household entrants on both cross sectional household and cross sectional individual weights for several panels that reflect the range of approaches:

- The Panel Study of Income Dynamics (PSID) began in 1968 as a representative sample of the US population and the households in which they reside. Just one person ("head of household") is interviewed per household. The PSID now covers roughly 9,000 households in the USA.

- The German Socio Economic Panel (SOEP) began in 1984. Every adult household member is sampled. SOEP has roughly 3300 responding households with 6000 responding persons.

- The British Household Survey Panel (BHPS) began in 1991. Every adult household member is sampled. The BHPS has roughly 4600 responding households with 8300 responding persons.

- The Swiss Household Panel (SHP) started in 1999. Every adult household member is sampled. SHP has roughly 7000 households with 18000 household members.

- The Household, Income and Labour Dynamics in Australia Survey (HILDA[2]) began in 2001. Every adult household member is sampled. HILDA has roughly 7200 responding households with 13300 responding persons.

[2] For convenience, we refer to the "the HILDA survey" simply as "HILDA".

Tables 1 and 2 show the effect of the approaches on individual and household weights, respectively. BHPS and SHP use the weight share method. SOEP and HILDA use the modeling approach. The PSID only interviews one household member thereby effectively assigning others the weight zero.

Table 1 further shows how individual weights are calculated from household weights and vice versa. For the two panels using the modeling approach (HILDA and SOEP), individual weights are derived from household weights. Because both panels select all adult household members, the selection probability of an individual is the same as the selection probability of a household. (In practice, due to individual nonresponse, individual weights may vary from household weights). The two panels using the shared weights approach (BHPS and SHP) compute the household weight as the average individual weight[3]. Because under fair shares all individuals receive the same weight, computing the household weight as the average individual weight or setting the household weight equal to the individual weight are equivalent.

For discussing the effect of household entrants on weights, we distinguish between regular household entrants, recent immigrants and births/adoptions.

The effect of regular household entrants on weights

When there are new household entrants, the individual weights and households weights of existing household members are down-weighted for both the modeling and the shared weights

3 For the BHPS, the average is computed over all household members, not just the wave-1 sample members (Nick Buck, personal communication).

approach. For the modeling approach, the household weight decreases because multiple paths of entry increase the selection probability of the household. For the shared weights approach, the individual weights decrease because the sum of the individual weights remains by definition constant. Therefore the BHPS household weight, the average individual weight of individual household members, also decreases.

From wave 2 onward, there are unknown selection probabilities (cf equation 1) for new household entrants. For both HILDA and SOEP, unknown selection probabilities are estimated via regression and used to compute the household weight. All individual weights are then derived from the (down-weighted) household weight, adjusted for attrition and post–stratification. Additional differences arise between HILDA and SOEP in their approach to modeling attrition. Briefly, HILDA models attrition from wave 1 to wave n rather than wave by wave like the SOEP.

Births and Adoptions

Births and adoptions (after wave 1) by definition could not have been sampled in wave 1. They represent the changing target population – the part of the population that did not exist in wave 1 - and are not treated like regular entrants. In the modeling approach, individual weights are typically set to the household weight. However, unlike for regular entrants, the household weight does not decrease. In the shared weights approach, births/ adoptions are also assigned additional weight. The BHPS assigns the average individual weight (not the shared weight) of the parents to births/ adoptions. If only one parent is a sample member, that child receives only half that weight (Taylor et al. 2009, p. A5-9). The PSID also assigns the average individual weights of the parents to births/ adoptions, unless only one parent lives in the household. In this case, birth/adoptions are assigned half that weight.

The SHP has not yet set rules for this issue because the children born into the panel are still too young to be interviewed. For the modeling approach, the household weights remain unchanged. For the BHPS and for the PSID, average individual weights are recomputed.[4]

"Understanding Society" (USoc) (http://www.iser.essex.ac.uk/survey/understanding-society), a large recent longitudinal panel of Great Britain (co-located with the BHPS whose sample became part of USoc), implements an alternative strategy of assigning weights to children. The expected number of children of two wave 1 respondents who marry spouses outside of the panel is twice as large as the expected number of children of two wave 1 parents. This may lead to an underrepresentation of children of wave 1 parents because wave 1 parents – already married in wave 1 – are on average older than parents in partnerships forming after wave 1. USoc assigns positive weight only to children where the mother was a wave 1 sample member, and zero weight to other children.[5]

[4] For the BHPS, a birth can lead to an increased household weight. Suppose there is a 3 person household: two wave 1 parents with weight 10 each and a grandmother who moved in after wave 1. A child is born and receives the average parent weight (10). The household weight before the birth was 20/3=6.7, the household weight after birth is 30/4=7.5.

[5] We thank Peter Lynn for pointing this out.

Recent Immigrants

Recent immigrants are individuals who have immigrated into the target population after wave 1 of the survey. They are not necessarily foreign nationals. Both recent immigrants and births represent groups of new entrants that could not have been sampled in wave 1. Because they change the target population, recent immigrants should be treated differently than other panel entrants. However, except for HILDA and SHP, panels treat immigrants just like other panel entrants.

In HILDA, when an immigrant joins a household, the household weight remains unchanged (for a regular entrant the household weight decreases). Therefore, individual weights of all household members are unchanged also (for a regular entrant, individual weights of all household members decrease) and, as with all other household members, the immigrant's weight equals the household weight. In SHP, when an immigrant joins a household, individual weights of existing members remain unchanged (for regular entrants, individual weights decrease). The recent immigrant is assigned the average weight of the original sample members in the household (Voorpostel et al. 2009, Section 4.2.3b) . The SHP defines the target population to exclude households composed exclusively of recent immigrants (Graf 2009, p.19) .

In SOEP, there is a special refresher sample just for immigrants. Recent immigrants (into existing households) outside this refresher sample are treated like any other household entrant. To the extent that panels like SOEP and BHPS do not treat recent immigrants differently from regular household entrants, we attribute this to the difficulty and the additional burden to distinguish between recent immigrants and regular household entrants.

Deaths and Emigration

Deaths and emigration are generally unproblematic from a survey perspective. In HILDA and SOEP, the weight of a dead person is simply removed because deaths change the size of the target population too. In the PSID the household weight computed as the average individual weight in the household has to be recalculated without the deceased person. Because in the BHPS the household weight is computed as an average of all members (including those with zero weight), the death of a member with zero weight increases the household weight and the death of a respondent with weight decreases the household weight. Individual weights do not change. The weight of a dead person is simply removed; it is not redistributed under "shared weights".

Household splits

A household split occurs when one or more members of a household leave a household (e.g. grown child, divorced spouse) and form a separate household. In the shared weight approach individual weights (not shared weights) remain with the individuals as they move to form new households. Respondents with a non-zero weight are wave 1 respondents and births/adoptions (and recent immigrants in the SHP). For example, suppose a wave 1 couple each with individual weight 10,000 separates, and the wife moves in with a new partner. Both respondents retain their individual weight of 10,000. The shared weight of the husband – now in a single household – remains 10,000 whereas the (fair shares) shared weight of the wife and her new partner is 5,000 each. The weight of all other respondents is zero and their zero weight is carried forward to new household. Therefore, the shared weight approach does not work well when such members are followed.

In the modeling approach all newly formed households get the same weight as the existing household. As before, individual weights are derived from household weights.

Household mergers

Our comparison revealed two different approaches under the label of "household merger". On closer inspection, they turned out to correspond to two types of household mergers: 1) unrelated merge: two or more unrelated sample households merge 2) move-back merge: two or more households re-merge after having formed a single household at some earlier time during the lifetime of the panel. For example, a grown child moves out of his parents home to go to college. After college, the grown child moves back in with his/her parents.

For the modeling approach, the unrelated merge is treated just like regular new household entrants with the one difference that the selection probabilities in equation (1) are known and need not be estimated. This type of merger is rare but has occurred in and is implemented in HILDA.[6]

The second type of merger is different in that the selection probability of a household does not change as the grown youngster moves back into the parent household. SOEP uses the former household weight corresponding to the new head of household (In SOEP, the head of household is the person who fills out the household questionnaire). This type of merge is also rare and has occurred less than 20 times in 26 SOEP waves.

---

[6] This type of merger has occurred in SOEP but is currently not treated as such.

For the shared weights approach this issues does not arise because the household weight is derived from individuals (rather than the other way around).

4. Conditions under which modeling and fair shares weights are identical

When sampling households and selecting all household members for inclusion in the panel, the individual weights of all household members are equal (before adjustments for non-response and post stratification). While unknown selection probabilities in equation (1) are estimated, the estimates serve to compute the household selection probability. Therefore, for the modeling approach respondents living in the same household have the same weight.

In general, the shared weights approach does not require equal weights. However, the most popular approach, "fair shares", implemented in the BHPS and in the SHP, does assume equal weights. We compute under which conditions the weights from the modeling approach and the "fair shares" approach coincide. The insight also points to one possible solution for the assignment of weights to respondents in households without a weight under the fair shares regime.

The modeling approach is equivalent to a fair shares approach if the sum of cross sectional individual weights do not change when one new entrant group moves into the household:

$$\sum_{i=1}^{n_1+n_2} w_{12} = \sum_{j=1}^{n_1} w_1$$

where n_1 is the number of individuals in the household before the arrival of a new entrant, w_1 is the individual weight beforehand and $w_{1,2}$ the corresponding weight afterwards. Weights are computed as inverse selection probabilities.

$$\sum_{i=1}^{n_1+n_2} 1/(p_1 + p_2 + p_3 - p_1 p_2) = \sum_{j=1}^{n_1} 1/p_1 \qquad (2)$$

where - like in equation (1) - p_1 is the selection probability of the existing household, and p_2 the selection probability of the new entrant. Solving for p_2 yields

$$p_2 = p_1 n_2 / n_1 \cdot 1/(1 - p_1) \qquad (3)$$

1/ (1-p_1) represents an adjustment factor which is close to 1.0 for small probabilities p_1 and equals 1.0 if the term p_1p_2 in equation (2) is removed. Therefore, the adjustment factor represents the probability of selecting both (rather than just one) constituent households in wave 1. Selection probabilities are typically very small. For SOEP, about 10,000 households are selected out of 40 million German households resulting in average selection probabilities in the order of 0.00025. For small probabilities, if one entrant (n_2=1) joins the household, the probability $p_2=p_1/n_1$. The probability for a new entrant is inversely proportional to the number of existing household members. Therefore, the sampling weight for a new entrant is proportional to the number of existing household members.

For example, suppose a single person moves into a two person household and that the household was selected in wave 1 with a probability p_1=0.01. This implies a weight of 100 for the household and each of the two persons (ignoring non-response and other adjustments). Using equation (3), the fair shares approach and the modeling approach yield the same sampling weights for the combined household if p_2= 0.005 (ignoring the higher order term p_1p_2). The individual (shared) sampling weight for each individual in the combined household is 200/3=66.7. Likewise, the individual weight for the modeling approach is 1/(0.01+0.005)=66.7. This is also the household weight for both approaches (for fair shares, the household weight is the average of 3 equal weights, see Table 1). While the two approaches give the same sampling weights for p_2=0.005, it is not clear how under what circumstances the selection probability of a one- person-household should be half that of the selection probability of a two-person-household.

The modeling approach coincides with the fair shares approach if a single new entrant is assigned the selection probability $p_1/n_1 * 1/(1-p_1)$. If the estimated probability for p_2 is larger, the modeling approach leads to a smaller weight than the fair shares approach and vice versa.

We will now look at some special cases. When the existing sample household consists of a single person (n=1) joined by a single other person (m=1, e.g. new partner moves in), equation (3) becomes

$$p_2 = p_1 /(1 - p_1) = odds(p_1)$$

Because the odds are always larger than the corresponding probability, this implies $p_2 > p_1$. For small selection probabilities, the probabilities are approximately equal: $p_1 \approx p_2$. For large selection probabilities p_1, the two approaches cannot coincide because the probability p_2 computed from (3) is greater than 1.0. As long as $p_1 < n_1 /(n_1 + n_2)$, we have $p_2<1$. Because selection probabilities are typically small this is not likely a problem in practice.

In the appendix we derive a formula for two entrant groups analogous to equation (3). Leaving out higher order terms in equation (1), we also develop an approximate formula for (k-1) entrant groups, and show that for two entrant groups the approximate and the exact formula give very similar results for small selection probabilities p_1 (Table A-1).

In summary, for small p_1 the modeling approach and the fair shares approach are approximately equivalent when the selection probability for a new entrant is inverse proportional to the number of existing household members. Differences related to households splits remain: When a household split occurs weights are redistributed under the fair shares approach but not for the modeling approach. This includes the case in which some respondents moving out are left without a weight in the faire shares approach ("orphan respondents").

5. Weights of orphan respondents in the shared weights approach

The shared weights approach is only fully appropriate as long as at least one person with a sampling weight (wave 1 members, births and recent immigrants) remains in the household (Lynn 2009, p.28). We call respondents in households without a person with a sampling weight "orphan respondents". If a panel using the shared weights approach chooses to follow respondents without sampling weights (e.g. spouses /partners who moved in after wave 1), those respondents cannot be assigned a shared weight when they move out (e.g. due to divorce /separation) by themselves[7]. This was not a problem up to know because the panels with the

[7] If they move out with a child that has a sampling weight, the shared weights approach works fine.

fair shares approaches (BHPS[8], SHP) did not follow household members without a weight as SOEP and HILDA do.

However, the Swiss household panel uses the fair shares approach for cross sectional weights (Graf 2009), and recently (Wave 9) changed the following rules to follow everyone (spouses, roommates, relatives, etc.)(Voorpostel et al. 2009, Section 2.3.2). This requires a revision of the approach to cross sectional weights. The option of assigning zero weight to orphan respondents is not appealing because it reduces the sample size.

A second option – not yet discussed in the literature – is to adopt a hybrid approach in which shared weights continue to be used and selection probabilities for new orphan respondents are estimated separately. Selection probabilities of orphan respondents need only be computed once when they first become orphan respondents. Subsequently, they are no longer orphan respondents and the shared weights approach can be applied.

The advantage of this hybrid approach is threefold: (1) it solves the problem of orphan respondents, (2) it allows for a smooth transition when following rules are expanded like in the Swiss household panel (as compared to switching completely to a modeling approach), and (3) potential bias and variability due to the model-based estimation are restricted to orphan respondents only.

A third option arises because weights computed under fair shares can be reinterpreted as being computed based on the modeling approach. Unlike the fair shares approach, the modeling

---

approach can assign weights to orphan respondents. This solution is not particularly attractive, though, because it is not clear how to justify the selection probability p_2 in equation (2) without the context of shared weights.

6. Discussion

We have discussed two common approaches to cross sectional weights in household panel surveys, “shared weights” and “modeling”. The fair shares approach is a modeling approach corresponding to a specific model of the selection probability of new entrants (equation 2). This is not true for the shared weights approach in general because individual shared weights do not have be equal; they just have to sum to the same constant. Specifically, the fair shares approach corresponds to a model in which the selection probability for a new entrant is (approximately) inverse proportional to the number of existing household members. That is, household entrants joining larger households correspond to smaller probabilities in the model. This model is not intuitive; we cannot think of a realistic sampling design for which this would be the case. Even though the estimates of weights coincide for both approaches, it is not clear how to justify equations (2), (A.2) and (A.3) from a modeling perspective.

The shared weights approach is limited in that it excludes orphan respondents, i.e. it requires the presence of one sample member with a weight in the household. This is problematic when the panel follows household members who moved in after wave 1 (such as spouses or partners) who later leave the household (e.g. divorce/separation). We have proposed a hybrid shared weights approach that models the selection probability of orphan respondents separately when they become orphans. While this appears to have some advantages, empirical work is needed to evaluate this procedure in practice.

The comparison of approaches to cross sectional weights has identified similarities and differences between the two approaches and between the panels. When regular new entrants join a household, the cross sectional weights of existing household members decrease for both approaches. The comparison has also uncovered two different types of household mergers which we have termed the “move-back merge” and the “unrelated merge”, the rare merge of two unrelated sample households. Some panels do not distinguish between recent (after wave 1) immigrants and regular household entrants. Presumably, the administrative burden of distinguishing between recent immigrants and regular respondents is high relative to the potential number of recent immigrants.

Appendix

In the main text we gave a formula with the conditions under which the modeling approach and the "fair shares" approach result in the same weights when one new entrant group enters a household (e.g. a mother and her child from a previous marriage move in with a respondent). Here we derive two additional formulas: (1) a formula for the case when there are *two* new entrant groups (e.g. a sample member and two college friends form a new household; the two college friends were previously living separately). (2) a approximation for an arbitrary number of new entrant groups when leaving out higher order terms in equation (1). For two entrant groups we have

$$\sum_{i=1}^{n_1+n_2+n_3} w_{123} = \sum_{j=1}^{n_1} w_1$$

This implies

$$\sum_{i=1}^{n_1+n_2+n_3} 1/(p_1 + p_2 + p_3 - p_1 p_2 - p_1 p_3 - p_2 p_3 + p_1 p_2 p_3) = \sum_{j=1}^{n_1} 1/p_1 \qquad \text{(A.1)}$$

Because there are two unknown probabilities, p_2 and p_3, but only one equation, there is no unique solution to the estimation. We make the simplifying assumption that the selection probabilities of new entrant groups are equal: $p_2 = p_3$. This assumption reflects a sampling design in which unobserved constituent households were selected with equal probability. Using this assumption solving for p_2 gives

$$p_2 = \left(1 - \sqrt{1 - \frac{p_1(n_2 + n_3)}{(1 - p_1)n_1}}\right) \qquad \text{(A.2)}$$

The second solution of the quadratic equation is not valid because it yields a probability greater than 1. Unfortunately, this equation is less interpretable than equation (3). A corresponding solution can be derived for 3 entrant groups which is even less interpretable. Instead, we try to gain intuition by developing an approximation removing the higher order terms in equation (A.1) and show that the two formulas give numerically similar results.

If we remove higher order terms in equation (A.1) which are very small, we have

$$\sum_{i=1}^{n_1+n_2+n_3} 1/(p_1 + p_2 + p_3 + \cdots + p_k) = \sum_{j=1}^{n} 1/p_1$$

Then,

$$p_2 = p_1 \frac{n_2 + n_3 + \cdots + n_k}{(k-1)n_1} \qquad \text{(A.3)}$$

If $n_2=\ldots=n_k=1$, this reduces to $p_2 = p_1/n_1$ or further to $p_2=p_1$ if $n_1=1$ also. Table A-1 compares the estimates of p_2 using the exact formula in (A.2) and the approximate formula without the higher order terms in (A.3) as a function of p_1. The estimate using the approximation is always smaller than the estimate using the exact formula. The approximation is not very good for larger values of p_1 (p_1=0.2 and p_1=0.1) but those values do not occur in practice. When the selection probability of the existing household, p_1, is 0.01 or less, the exact estimate of p_2 is only 1.5% larger than the approximate estimate. For small values of p_1, the higher order terms are negligible. Therefore, if higher order terms are negligible, the fair shares approach corresponds to a model in which selection probabilities are proportional to the number of people.

References

Ernst, LR. 1989. Weighting issues for longitudinal household and family estimates. In *Panel Surveys*, edited by D. Kasprzyk, G. Duncan, G. Kalton and S. M. P. New York: Wiley & Sons, Inc.

Galler, HP. 1987. Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In *Lebenslagen im Wandel: Analysen*, edited by H.-J. Krupp and U. Hanefeld. Frankfurt: Campus.

Graf, Eric. 2009. Weighting of the Swiss Household Panel: SHP I Wave 9, SHP II Wave 4, SHPI and SHP II combined: Swiss Foundation for Research in Social Sciences (FORS).

Horvitz, DG, and DJ Thompson. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47 (260):663-685.

Kalton, G, and JM Brick. 1994. Weighting schemes for household panel surveys. Paper read at Joint Statistical Meetings, Survey Research Methods Section.

Kalton, G, and JM Brick. 1995. Weighting schemes for household panel surveys. *Survey Methodology* 21 (2):33–34.

Lavallée, P. 1995. Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology* 21:25-32.

Lavallée, P, and P Caron. 2001. Estimation using the generalized weight share method: the case of record linkage. *Survey Methodology* 27 (2):155-169.

Lynn, P. 2009. *Methodology of Longitudinal Surveys*. Chichester: John Wiley & Sons Inc.

Rendtel, U, and T Harms. 2009. *Weighting and Calibration for Household Panels*. In *Methodology of longitudinal surveys*, edited by P. Lynn. Chichester: John Wiley & Sons Inc.

Taylor, MF, J Brice, N Buck, and E Prentice-Lane. 2009. British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex.

Voorpostel, M, R Tillmann, F Lebert, B Weaver, U Kuhn, O Lipps, V-A Ryser, F Schmid, and B Wernli. 2009. Swiss Household Panel User Guide (1999-2009):Wave 10: Swiss Foundation for Research in Social Sciences (FORS).

Watson, N. 2004. Wave 2 weighting. In *HILDA Project Technical Paper Series*. Melbourne: University of Melbourne.

Table 1: The effect of household changes on cross sectional individual weights for different household panels

	BHPS	SHP	PSID	HILDA	SOEP
Method for computing weights	**hh weight = average of individual weights**	**hh weight = average of individual weights**	**hh weight = average of individual weights**	**individual weights = household weight**	**individual weights = household weight**
Method for assigning weight to new Entrants	**Weight Share**	**Weight Share**	**Zero Weight**	**Modeling**	**Modeling**
Regular Household Entrants	down-weighted	down-weighted	zero weight	down-weighted	down-weighted
Immigrants	like other household entrants	average of (individual) OSM weights	like other household entrants	unchanged	like other household entrants
Birth / adoptions	receive average weight of parents	does not apply (panel is 11 yrs old, weights are assigned at age 14)	average weight of parents; if only one parent: 1/2 weight of head of household	receive household weight	receive household weight
Household Split	unchanged	zero in households without OSM, otherwise unchanged	unchanged	unchanged (splitting hhs receive the same hh weight)	unchanged (splitting hhs receive the same hh weight)
Merging households	unchanged	unchanged	unchanged	"unrelated merge": like regular household entrants	"move back merge": receive weight from new head of household
Death	unchanged for others	unchanged for others	unchanged for others	unchanged for others	unchanged for others

Table 2: The effect of household changes on household weights for each household panel. Notation: HH= household, OSM = Original sample member, TSM= Temporary sample member

	BHPS	SHP	PSID	HILDA	SOEP
Regular Household Entrants	down-weighted	down-weighted	unchanged (entrant is not OSM)	down-weighted to account for multiple pathways of being selected	down-weighted to account for multiple pathways of being selected
Births / adoptions	average is recomputed	does not apply (panel is 11 yrs old, weights will be assigned at age 14)	average is recomputed	unchanged	unchanged
Immigrants	treated like other household entrants	unchanged	treated like other household entrants	unchanged	treated like other household entrants
Household Split	average is recomputed	households without OSM: 0. Otherwise weight share	Averages are computed for each household separately	The same HH weight is carried over to both new households	The same HH weight is carried over to both new households
Merging households	Average is computed for merged household	Average is computed for merged household	Average is computed for merged household	"unrelated merge": like regular household entrants	"move-back merge": former household weight of the new head of household is used
Death	OSM death: down-weighted. TSM death: up-weighted	OSM death: down-weighted. TSM death: up-weighted	average is recomputed	unchanged	unchanged

Table A-1: Comparison between the exact formula for p_2 (equation A.2) and the approximate formula when higher order terms are omitted (equation A.3). Computations are based on 2 new entrant groups to a household. Each group consists of only one person (n_1=n_2=n_3=1). This table shows that for small values of p_1, the higher order terms are negligible.

p_1	**Exact p_2**	**Approx. p_2**	**Ratio (Approx. p_2 / Exact p_2)**
0.200000	0.292893	0.200000	1.464466
0.100000	0.118083	0.100000	1.180829
0.010000	0.010153	0.010000	1.015255
0.001000	0.001002	0.001000	1.001503
0.000100	0.000100	0.000100	1.000150
0.000010	0.000010	0.000010	1.000015
0.000001	0.000001	0.000001	1.000002