Deutsches Institut für
Wirtschaftsforschung

DIW BERLIN

www.diw.de

SOEP The German Socio-Economic Panel Study

SOEPpapers

on Multidisciplinary Panel Data Research

98

Martin Spiess • Martin Kroh • Rainer Pischner • Gert G. Wagner

On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP) —Tracing, Weighting, and Frequencies

Berlin, April 2008

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

Martin Spiess *
Martin Kroh **
Rainer Pischner ***
Gert G. Wagner****

On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP)—Tracing, Weighting, and Frequencies

Berlin, April 2008

* SOEP at DIW Berlin, and University of Flensburg; email: mspiess@diw.de

** SOEP at DIW Berlin; email: mkroh@diw.de

*** SOEP at DIW Berlin; email: rpischner@diw.de

*** SOEP at DIW Berlin, Berlin University of Technology (TUB), and Max Weber Center for Advanced Studies, Erfurt; email: gwagner@diw.de

Contents

1 Introduction

Prospective household panel studies like PSID (USA) and the German SOEP were originally designed to give a picture of the current and future composition of private households and the well-being of the members of those households (e.g., Wagner et al. 1993, 2007). Over time, however, it has become more and more obvious that the individual-level data from household panel surveys can be an excellent source for the analysis of "intergenerational transmissions"; e.g., the transmission of poverty from parents to children (Jenkins and Siedler 2007), the well-being of widows and widowers (Burkhauser et al. 2005, Lucas 2007), and the impact of children's well-being on the happiness of their parents (Schwarze and Winkelmann 2005). Further exciting research questions have recently been opened up by the possibilities of linked data—not only on parents and children but also on the linked life courses, e.g., of siblings and couples (Ermish et al. 2006). Linked life courses allow analyses along the lines of the "behavioural genetics" approach, which attempts to disentangle the impacts of nature ("genes") and social circumstances ("environment") on human behaviour and well-being. Schimmack and Lucas (2007), for example, used SOEP data to analyze the well-being of the same couples during marriage and after divorce.

Whereas all household panel studies in developed countries provide data on parents, children, and siblings, there exists only one study to date that provides data on both ex-partners after a divorce or split in a cohabiting unit: namely, the SOEP (Jenkins and Siedler 2007, 6). This is due to the tracing rules adopted for the SOEP, in which all household members are traced. In particular, if an individual enters the sample after the corresponding household itself joined the survey, he or she will continue to be surveyed even after leaving the sample household.

In this paper we discuss the rationale for tracing non-original sample members (Non-OSMs) in household panel studies, and in particular in SOEP, and the implications for weighting (Section 2 to Section 5). In Section 6 and the Appendix we present results on the incidence, survival rates, and thus the relevance of Non-OSMs in the SOEP.

2 Tracing Rules in Household Panel Studies

Because households are "dynamic units," the question of who should be a sample member in longitudinal *household* panels is not as trivial as it may at first seem. Whereas this question is easy to answer in the case of cross-sections and even cohort studies of persons, in household

panels where the household is the selected unit, the answer becomes much more difficult to answer due to births, move-ins, and split-offs.

From a cross-sectional perspective, additional difficulties arise in household panel surveys through immigrants, who form new households and thus become relevant population units. In most panel studies (e.g., BHPS and PSID), these units are not covered, but the German SOEP accounts even for immigrants that do not enter existing households through its special "immigrant sample" (cf. Burkhauser et al. 1997) and through general refreshment samples (cf. Wagner et al. 2007). However, we will not pursue this topic further in the present paper.

Household Panel Studies like the BHPS, the PSID, and the European Community Household Panel (ECHP) do not follow Non-OSMs once they leave households if OSMs stay. However, there are exceptions. BHPS, for example, follows the Non-OSM parents of OSM children. However, if parenthood is deemed to constitute an "important relation" to an OSM in the BHPS (Jenkins and Siedler 2007, 2006), leading to the tracing of Non-OSMs, then one must ask why a former marriage or partnership of an OSM to a Non-OSM is not sufficiently "important". Furthermore, the living conditions, income, and well-being of widows/widowers who are Non-OSMs[1] may be of interest as well. Every widow/widower, whether an OSM or not, can contribute to this kind of research. In addition, widows/widowers who are Non-OSMs can provide very valuable information about the "final resting place" of the deceased OSM (cf. Gerstorf et al. 2007). In the SOEP study, a special questionnaire about the final resting place and bequests ("exit interview") has been developed, which we will test in 2008.

3 Tracing Rules of SOEP

SOEP was started with the "classical" tracing rule that only OSMs should be traced. The set of OSMs includes all respondents of wave one and their children living in the same wave-one-household. However, two groups of Non-OSM not covered by this "cross-sectionally"-based tracing rule are unborn children and children living abroad. But if they later become members of the relevant population, they should be sampled once they appear in the sampling area, i.e., in private households in Germany. And in fact, in all household panels, these children ("virtual OSMs") do become OSMs once they show up in their parents' households. In

[1] In the very first waves of a household panel study, all widows/widowers are OSMs because they are living in the households originally sampled. Over the course of time, however, the percentage of Non-OSM widows/widowers in the sample increases because more and more widows/widowers enter the sample through marriage to OSMs.

the first few waves of SOEP as in other panel studies, other Non-OSMs are only interviewed as long as they live together with OSMs in a sample household. Interviewing Non-OSMs during that time is necessary in order to obtain a full picture of the household, in the SOEP especially the household income, an OSM is living in.

The SOEP fieldwork followed this "classic concept" for the first six waves (1984-89). However, the experience was that interviewers were often not able to distinguish between OSMs and Non-OSMs. This resulted in a substantial portion of Non-OSM persons being interviewed accidentally. Yet, this data turned out to be of particular interest for substantive studies analyzing, for example, the consequences of divorce.

Beginning with wave seven (1990), the decision was made by the SOEP group in Berlin and the fieldwork organization Infratest Sozialforschung (Munich) to follow all persons that had ever been interviewed once.[2] In principle, this tracing rule creates a kind of snowball-effect that would theoretically include—in the very long run—the entire population of Germany. However, this does not happen due to the attrition rates of households and individuals. Equally important is that—by lucky accident—the weighting concept of SOEP was designed in a manner that allowed the weighting of Non-OSM to be dealt with appropriately.[3]

4 Weighting Scheme of SOEP: Fusion of OSM and Non-OSM Households

The basic idea of the SOEP weighting scheme is that the reciprocals of the weights can be interpreted as the (estimated) probabilities of observing the corresponding units (cf. Galler, 1987; Rendtel 1995). This idea is in line with design-based as well as modern model-based approaches to compensate for different sampling and response probabilities (e.g., Robins, Rotnitzky and Zhao, 1995; Särndal, Swensson and Wretman, 1992; Wooldridge, 2002).

At the core of the weighting scheme is the selection probability of each unit given by the sampling design and the (estimated) probability of observing the units selected into the sample. From a conceptual point of view, deriving the weights causes no problem in the first wave of a panel. Since the definition of the sampling units, i.e., households, is unambiguous in this case, the problem comes down to finding variables that allow the prediction of re-

[2] To the best of our recollection, this was proposed by the head of the fieldwork organization, Bernhard von Rosenbladt.

[3] See Heinz P. Galler (1987), to whom we are deeply indebted for his highly flexible weighting concept.

sponse given selection into the sample in such a way as to allow consistent estimation, preferably for a wide range of substantive analyses. Conceptual problems may arise, however, in the longitudinal setting, i.e., from wave two on. Then, since households are artificial units and one household can split up into two or even more households, or two or more households can fuse into one from one wave to the next, the derivation of the probabilities of the artificial units is not straightforward and depends on an arbitrary definition of a sample member. In this section, we will concentrate on the derivation of the weights for those households in which a Non-OSM moves in between wave $t-1$ and the current wave t.

There are at least three arguments why non-sample-individuals should be included in the sample and the weighting scheme once they enter a sample household. First, there are substantive considerations as explained in the introduction above, i.e. if one is interested in the linkage of life trajectories (couples, parents and children). Second, if Non-OSMs are systematically different from OSMs, e.g. in terms of mobility, ignoring these cases may lead to an underestimation of the dynamics of interpersonal relations over the life course. Third, ignoring these individuals may lead to conceptual and empirical problems that are probably more serious than if they were included in the sample and the weighting scheme. For example, if a sample household is inhabited by a couple consisting of a Non-OSM and an OSM and their common children, then it is not straightforward to justify why these children should be Non-OSMs or OSMs.

The core of the problem of including Non-OSMs in the weighting scheme as soon as they move into a sample household is how to derive the probability of observing the unit (household) in the actual wave at time t given a fusion of two (or more) households from wave $t-1$ to wave t. Since the most usual incidence is the fusion of two households, this is the case considered in this paper. Three possibilities must be distinguished, where we use the terms FWSH (former wave sample household) and Non-FWSH (non-former wave sample household) as referring to their state in $t-1$ as being a sample unit or not (for details see Rendtel 1995):

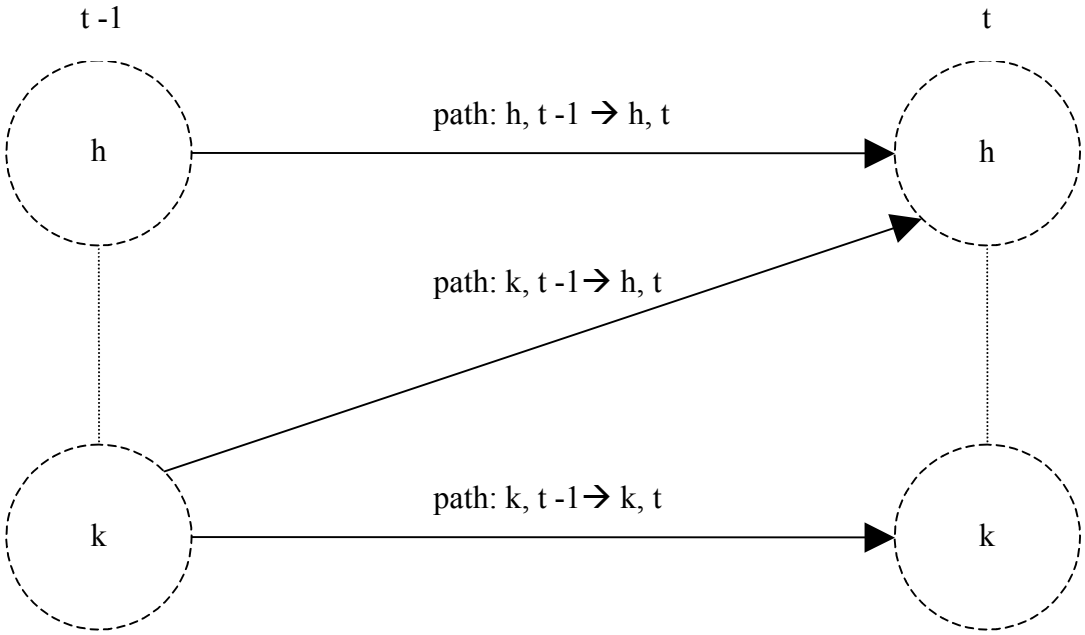
1) The two households are FWSHs,
2) one household is an FWSH, the other is a Non-FWSH but belongs to the population of interest, and,
3) one household is an FWSH, the other is a Non-FWSH and does not belong to the population of interest.

Given a panel data set and assuming a fixed population from $t-1$ to *t*, the probability of observing a household, say *h*, in wave $t>1$ that is a fusion of two households is given by the probability of observing *h* starting in $t-1$ from household *h* plus the probability of observing *h* starting from the other household, say *k,* in $t-1$ from household *k*, minus the probability of observing *h* in *t* starting from *h* and *k* in $t-1$. If $\pi_{h,t|k,t-1}$ denotes the probability of observing household *h* in t given household k was observed in $t-1$, then the probability of observing household *h* in *t*, given it was observed in $t-1$, can be written as

$$\pi_{h,t} = \pi_{h,t-1}\pi_{h,t|h,t-1} + \pi_{k,t-1}\pi_{h,t|k,t-1} - \pi_{h,k,t-1}\pi_{h,t|h,k,t-1} \qquad (1),$$

where $\pi_{h,t}$ denotes the probability of observing household *h* in t, $\pi_{h,t|h,k,t-1}$ denotes the probability of observing *h* via both paths, from *h* and k in $t-1$, and $\pi_{h,k,t-1}$ denotes the probability of observing both households in $t-1$. The situation is illustrated in Figure 1.

Figure 1: The follow-up paths of SOEP households from *t-1* to *t*.



Returning to the three possibilities of a fusion of two households, the third case, in which a household moves from abroad to Germany (the population of interest) and merges with an existing SOEP-household, is the simplest to deal with: the household that does not belong to

the population in $t-1$ has observation probability zero, and thus household *h* can only be reached via the path from *h* in $t-1$ to *h* in *t*, i.e.

$$\pi_{h,t} = \pi_{h,t-1}\pi_{h,t|h,t-1}\,.$$

If both households are sample households in $t-1$ (case 1), then the probability is as given in (1). However, since the last term in (1) is, in most cases, very small as compared to the other probabilities, it is ignored, and the probability of observing *h* in *t* reduces to

$$\pi_{h,t} = \pi_{h,t-1}\pi_{h,t|h,t-1} + \pi_{k,t-1}\pi_{h,t|k,t-1}\,.$$

All the probabilities involved can be derived or, under assumptions, be estimated.

Given the assumption that the third term in (1) can be ignored, case 2 is still not straightforward, as the corresponding variables at the household level for household *k* in $t-1$ are not observed. In particular, without several model assumptions, the probability $\pi_{k,t-1}$ cannot be estimated based on household information from $t-1$. One way to tackle this problem is to estimate $\pi_{k,t-1}$ based on individual information, i.e., from information given by the originally non-sample individual moving into the sample household ("new sample members"). If this individual refuses to participate, then even this information is not available. Then, since this refusal can be interpreted as the result of a strong nonresponse tendency (note that other members of the same household do participate), the probability $\pi_{k,t-1}$ is set equal to zero.

However, if the former non-sample individual participates, then we estimate $\pi_{k,t-1}$ following Rendtel (1995):

First, a linear regression model

$$y_{i,t-1} \equiv \log\left(\frac{\pi_{i,t-1}}{1-\pi_{i,t-1}}\right) = \alpha + x'_{i,t-1}\beta + \varepsilon_i \qquad (2)$$

is fitted, where $\pi_{i,t-1}$ is the probability of observing individual *i* in $t-1$, α is a constant, x_i is a vector of observed individual, preferable time-invariant characteristics, and ε_i is an individual error term. Note that this model is estimated based on the *observed response*, $y_{i,t-1}$, of *sample individuals* of all sample households in $t-1$. Second, under the *assumption* that the model approximately holds for the corresponding probabilities in $t-1$ for the new sample

members t', we can estimate the probabilities $\pi_{i',t-1}$ by first predicting the logits for the new sample members, $\hat{y}_{i',t-1}$, based on the same variables and parameter estimates from model (2) and then transforming them into the estimated probabilities

$$\hat{\pi}_{i',t-1} = \frac{\exp(\hat{y}_{i',t-1})}{1+\exp(\hat{y}_{i',t-1})}.$$

An *additional underlying assumption* is that we can estimate the probabilities $\hat{\pi}_{i',t-1}$ by $\hat{\pi}_{i',t-1}$ based on individual information observed at time *t*. An example from the 2006 wave is given in the next section.

Since there is no information available to estimate the probability $\pi_{h,t|k,t-1}$, we need a further restriction to be able to proceed. Thus, we *assume* that $\pi_{h,t|k,t-1} \approx \pi_{h,t|h,t-1}$. The actual probability of observing sample household *h* in *t* with new sample members moving in from $t-1$ to *t* can then be estimated by

$$\hat{\pi}_{h,t} = (\hat{\pi}_{h,t-1} + \hat{\pi}_{k,t-1})\hat{\pi}_{h,t|h,t-1}.$$

The weights, which are the starting point for the estimation of the final weights as delivered with the SOEP (e.g., Spiess & Kroh 2008; Haisken-DeNew and Frick, 2005), are then equal to the inverse of the estimated observation probabilities $\hat{\pi}_{h,t}$.

According to the strategy described above and the tracing rules adopted for the SOEP, each Non-OSM receives a weight once s/he enters a sample household which is a kind of starting weight for Non-OSMs. Note that the Non-OSM remains a sample member even if s/he leaves the sample household. *Further,* note that if a non-OSM would not be traced after leaving the sample household in a later wave, which simply means discarding the unit from the sample, then this would imply *missingness by design*.

5 Deriving Weights for the 2006 Wave

To give an example of how the weights for those households with non-sample move-ins between $t-1$ and *t* are estimated, we describe the corresponding prediction model for the 2006 wave. Note that because we are interested in predicting the logits (probabilities) for non-

sample individuals in $t-1$, only those parameter estimates are included in the prediction models that are significantly different from zero. The model estimated in 2006 based on observed *t-1* information (see equation (2) in Section 3) is based on 20,751 individuals. The covariates included and the estimated regression parameters (significantly different from zero at the $\alpha = 0.01$-level) are given in Table 1.

Table 1 Regression Model (Logit) of Individual Participation in 2005

	COEFFICIENTS
Intercept	-8.555
Unmarried	-0.163
Number of children in household	0.116
Age (*Reference: <46*)	
46-65	-0.163
>65	-0.148
Subsample (*Reference: Subsample A*)	
Subsample B (Turkish)	0.775
Subsample B (Yugoslav)	0.944
Subsample B (Greek)	1.483
Subsample B (Italian)	0.956
Subsample B (Spanish)	1.894
Subsample C (East Germans)	0.552
Subsample D (Immigrants)	not sign
Subsample E (Refreshment)	-0.474
Subsample F (Refreshment)	-0.736
Subsample G (High Income)	-0.219
Immigrant (*Reference: Native*)	
Ethic German immigrant	0.402
Other immigrants	-0.423
Place of origin (*Reference: not applicable*)	
West Germany	0.504
East Germany	0.612
Foreign Country	0.493

Table 1 proceeds on next page

Table 1 continued

Gender (*Reference: male*)	
Female	-0.045
School degree (*Reference: not applicable*)	
Secondary School Degree	0.326
Intermediate School Degree	0.285
Technical/Upper Second. Degree	0.250
Other Degree	0.223
Dropout, No School Degree	0.297
Interaction-Terms	
German nationality * Subsample F	0.529
Age 66 and older * Female	-0.123
Subsample D1 * No of kids	-0.150
Subsample D2 * No of kids	-0.099
Subsample G * Unmarried	0.330
Subsample G * Age 66 and older	0.337
Subsample G * Age between 46 and 66	0.287
Subsample G * No of kids	-0.087
Subsample G * Residence from former West-Germany	-0.166
Subsample G * Female	0.094
Subsample G * Technical/Upper Secondary Degree	0.114

6 Frequency and Participatory Behavior of Non-OSMs

This section provides some descriptive figures on the growing share of non-original sample members in the SOEP. Moreover, the section presents some evidence on the likelihood of refusals by new sample members as opposed to original sample members.

Table 2 reports the frequency and the relative share of households in 2006 by their composition of individuals who were either members of the originally sampled households or were not part of the initially sampled SOEP households. This leads to three types of households: those populated exclusively by individuals who belong to the original sample of households (here:

OSM-HH), those populated both by individuals from original sample households *and* by individuals from non-sampled households (here: Mixed HH), and lastly, households populated only by individuals from originally non-sampled households (here: Non-OSM HH).

Table 2 The Number of Households in SOEP 2006 by Sample and Non/OSM Status.

Samples	Number of Households				Shares of OSM Status in Percent		
	Total	OSM HH*)	Mixed HH	Non-OSM HH**)	OSM HH*)	Mixed HH	Non-OSM HH**)
A	2821	1572	950	299	0.557	0.337	0.106
B	655	392	223	40	0.599	0.340	0.061
C	1717	1123	461	133	0.654	0.268	0.078
D	222	150	68	4	0.676	0.306	0.018
E	686	567	96	23	0.826	0.140	0.034
F	3895	3394	450	51	0.871	0.116	0.013
G	859	786	69	4	0.915	0.080	0.005
H	1506	1506	-	-	1.000	1.000	-
All	12361	9490	2317	554	0.768	0.187	0.045

Note. *) OSM-HHs are households with original sample members only. **) Non-OSM-HHs are households with non-original sample members only. *Source*. SOEP (Waves A to W).

Among the 12,361 households surveyed in 2006, more than 20% contain at least a single person not covered by the originally sampled households. This share steadily increases as a function of the age of subsamples: the most recent subsample H from 2006, by definition, includes only originally sampled persons in each interviewed household. In subsamples F and G, drawn in 2000 and 2002, the OSM-HH reach a share of only 90% after seven and five waves, respectively. In the oldest subsamples, A and B, slightly more than half of the households were populated exclusively by respondents who belonged as members, children living abroad, or as unborn children to the household drawn in 1984. Interestingly, more than 10% of the households in subsample A of 2006 contain *no* individual who was part of the original sample of households. In the Appendix, we report the development of these different types of households in subsamples A through G separately.

Table 2 suggests—as one would expect—an increasing weight of households including (only) new sample members. An important question for the continuity of a long-running panel such as the SOEP is whether this population differs in its response behaviour compared to the original sample members. Are they, for instance, more difficult to follow up on, or are they more likely to refuse participation? As the question of long-term participatory behaviour is difficult to address at the household level (since households can, in principle, switch status repeatedly between the OSM, Mixed, and Non-OSM types), we investigate the participatory behaviour of individual respondents. Table 3 distinguishes three groups of respondents. The first group is made up by those who were members of the originally sampled households *and* were interviewed in the initial wave of each subsample A through G in 1984, 1990, 1994/5, 1998, 2000, and 2002, $t_{0,A\text{-}G}$. The second group contains individuals who were members of the originally sampled households in subsamples A through G but were *not* interviewed in $t_{0,A\text{-}G}$. This applies, for instance, to individuals who were too young to participate in $t_{0,A\text{-}G}$ and became part of the active sample in one of the following years $t_{>0,A\text{-}G}$. The final group contains respondents who were not members of sampled households in $t_{0,A\text{-}G}$ and thus participated, like second group, for the first time in $t_{>0,A\text{-}G}$, i.e., a wave subsequent to initial sampling. While the first two groups represent the raw sample of individuals already living in the original sample of households, the third group includes external entrants to the survey.

Table 3 The Probability of Continued Participation of Persons by Non/OSM-Status.

Years After First Interview	Respondents from Original Gross Samples in t_0		Respondents Entering the Gross Sample after t_0
	Participants in t_0	Non-Participants in t_0	
1	0.881	0.918	0.912
2	0.814	0.853	0.844
3	0.767	0.793	0.788
4	0.721	0.743	0.744
5	0.685	0.688	0.704
(...)			
10	0.563	0.477	0.530
(...)			
15	0.461	0.371	0.412
(...)			
20	0.379	0.274	0.311
N	35899	5268	6275
Mean Age in t_0	44.75	19.03	29.60

Note. Entries denote Kaplan-Meier survival estimates of individual respondents' participation in the SOEP after their first interview. If respondents move abroad or die, we consider this event as a form of right-censoring. *Source*. SOEP (Waves A to W).

Table 3 reports the probability of continued participation in the SOEP after each individual's first interview. Note that this time point coincides with the year of the first waves of subsamples A through G only for the first group. Note also that the reported Kaplan-Meier survival estimates treat an exit from the survey due to moving abroad and death technically as a form of right-censoring which does not affect the estimate of the probability of continued participation. The figures suggest that until the 6^{th} wave of each individual's initial interview, new sample members have, with 70%, an even somewhat *higher* response probability than interviewees who already lived in the originally sampled households at $t_{0,A\text{-}G}$ with 69%. Only in the very long run is the continued participation of initial first-wave respondents better than in the two other groups. The latter may be due to the much low(er) age of the respondents who enter the SOEP in a wave subsequent to initial sampling (mean age in t_0 of 19 and 30 respectively) as opposed to the sample of participating respondents in the initial wave (mean age in $t_{0,A\text{-}G}$ of 45 years). There is, however, no indication that new sample members are distinctly more volatile respondents than those in the original sample of households.

7 Conclusions

The German Socio-Economic Panel Study (SOEP) is a household panel survey with a different tracing rule than other household panel studies like BHPS or PSID. Whereas BHPS in principle and PSID drop Non-OSMs once the OSM leaves a sample household, SOEP traces these individuals. That is, even if a new (Non-OSM) household member leaves the sample household, the SOEP considers this individual an established part of the survey and continues to trace his or her subsequent living arrangements. This strategy, originally adopted on a non-theoretically basis which made fieldwork more efficient, turned out to enrich the data significantly.

Non-OSMs allow researchers to address innovative research questions. It has been clear since the very beginning that following Non-OSMs can be helpful in analyzing the impact of events like divorces or separations of cohabiting units. However, it has just recently been shown that the increased number of cases providing data on respondents who have lived together for some time and split up is extremely valuable for disentangling the influence of genes and environment based on the differences in biological and (changing) social factors (cf. Schimmack and Lucas 2007). In addition, the tracing Non-OSMs is helpful for the analysis of the terminal phase of life (cf. Gerstorf et al. 2008). The value of this tracing rule will increase further after the introduction of "exit interviews" into the SOEP regarding the terminal phase of life, death, and bequests of respondents who have passed away.

Non-OSMs eventually make up a large portion of the respondents and households in any long-running panel design. Preliminary analyses reported in this paper do not suggest that these cases are more volatile in their participatory behaviour than OSMs. Furthermore, excluding Non-OSMs once they have left a sample household may lead to conceptual and methodological difficulties.

8 References

Burkhauser, Richard V., Philip *Giles,* Dean R. *Lillard, and* Johannes *Schwarze* (2005): Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. *Journal of Gerontology* - Series B: Social Sciences. 60 (5): 238-246.

Burkhauser, Richard V., Michaela Kreyenfeld and Gert G. Wagner (1997): The German Socio-Economic Panel: A representative sample of reunited Germany and its parts. *Vierteljahrshefte zur Wirtschaftsforschung* 66 (1): 7-15.

Ermish, John, Marco *Francesconi,* and Thomas *Siedler* (2006): Intergenerational Mobility and Martial Sorting. *The Economic Journal* 116: 659-679.

Galler, Heinz P. (1987): Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In: *Hanefeld*, Ute and Hans-Jürgen *Krupp* (eds.), *Lebenslagen im Wandel*. Frankfurt and New York, 295-317.

Gerstorf, Denis., Nilam *Ram*, Ryne *Estabrook,* Jürgen *Schupp*, Gert G. *Wagner* and Ulman *Lindenberger* (2008): Life Satisfaction Shows Terminal Decline in Old Age: Longitudinal Evidence from the German Socioeconomic Panel Study. *Developmental Psychology* (in press).

Haisken-DeNew, John P., and *Frick*, Joachim R. (2005). DTC Desktop Companion to the German Socio-Economic Panel (SOEP). Version 8.0 - Dec 2005, Updated to Wave 21 (U). http://www.diw.de/english/soep/services__documentation/27905.html

Jenkins, Stephen P. and Thomas *Siedler* (2007): Using household panel data to understand the intergenerational transmission of poverty. DIW Discussion Paper No. 694. Berlin.

Lucas, Richard A. (2007): Adaption and the Set-Point Model of Subjective Well-Being. *Current Directions in Psychological Science* 16 (2): 75-79.

Rendtel, Ulrich (1995). *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität*. Frankfurt/New York: Campus.

Robins, James M., Andrea *Rotnitzky*, and Lue P. *Zhao* (1995): Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association* 90(429), 106-121.

Särndal, Carl-Erik, Bengt *Swensson* and Jan *Wretman* (1992): *Model Assisted Survey Sampling*. New York: Springer.

Schimmack, Ulrich and Richard *Lucas* (2007): Marriage Matters: Spousal Similarity in Life Satisfaction. *Schmollers Jahrbuch* 127 (1), 105-111.

Schwarze, Johannes and Rainer *Winkelmann* (2005): What can happiness research tell us about altruism? Evidence from the German Socio-Economic Panel. DIW Discussion Paper No. 475.

Spiess, Martin and Martin *Kroh* (2008): Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2006). DIW Data Documentation No. 27.

Wagner, Gert G., Richard V. *Burkhauser,* and Friedericke *Behringer* (1993): The English Language Public Use File of the German Socio-Economic Panel. *The Journal of Human Resources* 28 (2), 429-433.

Wagner, Gert G., Joachim R. *Frick*, and Jürgen *Schupp* (2007): The German Socio-Economic Panel Study (SOEP)—Scope, Evolution and Enhancements. *Schmollers Jahrbuch* 127 (1), 139-169.

Wooldridge, Jeffrey M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Appendix

Table A1-A

Status of SOEP Households in Sample A

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
1984(A)	4,528	4,528	-	-	100.0	-	-
1985(B)	4,141	3,968	173	-	95.8	4.2	-
1986(C)	3,962	3,643	319	-	91.9	8.1	-
1987(D)	3,910	3,509	401	-	89.7	10.3	-
1988(E)	3,743	3,279	463	1	87.6	12.4	0.0
1989(F)	3,647	3,093	552	2	84.8	15.1	0.1
1990(G)	3,612	2,933	654	25	81.2	18.1	0.7
1991(H)	3,613	2,833	738	42	78.4	20.4	1.2
1992(I)	3,585	2,723	798	64	76.0	22.3	1.8
1993(J)	3,603	2,657	842	104	73.7	23.4	2.9
1994(K)	3,577	2,541	896	140	71.0	25.1	3,9
1995(L)	3,526	2,417	945	164	68.6	26.8	4.6
1996(M)	3,485	2,333	967	185	66.9	27.8	5.3
1997(N)	3,458	2,240	1,011	207	64.8	29.2	6.0
1998(O)	3,387	2,154	1,016	217	63.6	30.0	6.4
1999(P)	3,325	2,055	1,040	230	61.8	31.3	6.9
2000(Q)	3,240	1,984	1,016	240	61.2	31.4	7.4
2001(R)	3,168	1,896	1,015	257	59.9	32.0	8.1
2002(S)	3,123	1,847	1,010	266	59.1	32.3	8.5
2003(T)	3,072	1,784	1,009	279	58.1	32.8	9.1
2004(U)	3,010	1,718	1,004	288	57.1	33.4	9.6
2005(V)	2,937	1,655	977	305	56.3	33.3	10.4
2006(W)	2,821	1,572	950	299	55.7	33.7	10.6
Total(A to W)	80,473	59,362	17,796	3,315	73.8	22.1	4.1

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves A to W); author's calculations.

Table A1-B

Status of SOEP Households in Sample B

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
1984(A)	1,393	1,393	-	-	100.0	-	-
1985(B)	1,181	1,170	11	-	99.1	0.9	-
1986(C)	1,128	1,085	43	-	96.2	3.8	-
1987(D)	1,116	1,030	86	-	92.3	7.7	-
1988(E)	1,071	952	119	-	88.9	11.1	-
1989(F)	1,043	889	152	2	85.2	14.6	0.2
1990(G)	1,028	852	167	9	82.9	16.2	0.9
1991(H)	1,056	848	189	19	80.3	17.9	1.8
1992(I)	1,060	828	212	20	78.1	20.0	1.9
1993(J)	1,064	814	227	23	76.5	21.3	2.2
1994(K)	1,023	763	233	27	74.6	22.8	2.6
1995(L)	982	717	238	27	73.0	24.2	2.8
1996(M)	960	676	252	32	70.4	26.3	3.3
1997(N)	931	637	259	35	68.4	27.8	3.8
1998(O)	898	607	249	42	67.6	27.7	4.7
1999(P)	858	570	246	42	66.4	28.7	4.9
2000(Q)	820	532	249	39	64.9	30.4	4.8
2001(R)	809	519	250	40	64.2	30.9	4.9
2002(S)	766	483	244	39	63.1	31.9	5.1
2003(T)	742	462	237	43	62.3	31.9	5.8
2004(U)	714	431	240	43	60.4	33.6	6.0
2005(V)	698	418	237	43	59.9	34.0	6.2
2006(W)	655	392	223	40	59.8	34.1	6.1
Total(A to W)	21,996	17,068	4,363	565	77.6	19.8	2.6

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves A to W); author's calculations.

Table A1-C

Status of SOEP Households in Sample C

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
1990(G)	2,179	2,179	-	-	100.0	-	-
1991(H)	2,030	1,968	61	1	97.0	3.0	0.0
1992(I)	2,020	1,888	125	7	93.5	6.2	0.3
1993(J)	1,970	1,792	167	11	90.1	8.5	0.6
1994(K)	1,959	1,701	237	21	86.8	12.1	1.1
1995(L)	1,938	1,635	272	31	84.4	14.0	1.6
1996(M)	1,951	1,602	311	38	82.1	15.9	2.0
1997(N)	1,942	1,549	339	54	79.8	17.	2.8
1998(O)	1,886	1,485	345	56	78.4	18.3	3.0
1999(P)	1,894	1,458	366	70	77.0	19.3	3.7
2000(Q)	1,879	1,409	397	73	75.0	21.1	3.9
2001(R)	1,850	1,367	399	84	73.9	21.6	4.5
2002(S)	1,818	1,321	404	93	72.7	22.2	5.1
2003(T)	1,807	1,256	437	114	69.5	24.2	6.3
2004(U)	1,813	1,209	479	125	66.7	26.4	6.7
2005(V)	1,771	1,182	460	129	66.7	26.0	7.3
2006(W)	1,717	1,123	461	133	65.4	26.8	7.8
Total(G to W)	32,424	26,124	5,260	1,040	80.6	16.2	3.2

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves G to W); author's calculations.

Table A1-D

Status of SOEP Households in Sample D

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
1995(L)	322	316	6	-	98.1	1.9	-
1996(M)	302	287	15	-	95.0	5.0	-
1997(N)	286	259	27	-	90.6	9.4	-
1998(O)	259	224	35	-	86.5	13.5	-
1999(P)	252	202	49	1	80.2	19.4	0.4
2000(Q)	249	197	48	4	79.1	19.3	1.6
2001(R)	234	182	51	1	77.8	21.8	0.4
2002(S)	244	177	64	3	72.5	26.2	1.2
2003(T)	248	176	67	5	71.0	27.0	2.0
2004(U)	236	165	66	5	69.9	28.0	2.1
2005(V)	233	155	75	3	66.5	32.2	1.3
2006(W)	222	150	68	4	67.6	30.6	1.8
Total(L to W)	3,087	2,490	571	26	80.7	18.5	0.8

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves L to W); author's calculations.

Table A1-E

Status of SOEP Households in Sample E

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
1998(O)	1056	1056	-	-	100.0	-	-
1999(P)	886	862	24	-	97.3	2.7	-
2000(Q)	838	793	43	2	94.6	5.1	0.2
2001(R)	811	745	60	6	91.9	7.4	0.7
2002(S)	773	689	73	11	89.1	9.4	1.4
2003(T)	744	646	83	15	86.8	11.2	2.0
2004(U)	732	623	93	16	85.1	12.7	2.2
2005(V)	706	593	95	18	84.0	13.5	2.6
2006(W)	686	567	96	23	82.7	14.0	3.4
Total(O to W)	7232	6574	567	91	90.9	7.8	1.3

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves O to W); author's calculations.

Table A1-F

Status of SOEP Households in Sample F

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
2000(Q)	6,052	6,052	-	-	100.0	-	-
2001(R)	4,911	4,796	115	-	97.7	2.3	-
2002(S)	4,586	4,380	200	6	95.5	4.4	0.1
2003(T)	4,386	4,081	295	10	93.1	6.7	0.2
2004(U)	4,235	3,836	373	26	90.6	8.8	0.6
2005(V)	4,070	3,613	415	42	88.8	10.2	1.0
2006(W)	3,895	3,394	450	51	87.1	11.6	1.3
Total(Q to W)	32,135	30,152	1,848	135	93.8	5.8	0.4

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves Q to W); author's calculations.

Table A1-G

Status of SOEP Households in Sample G

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
2002(S)	998	998	-	-	100.0	-	-
2003(T)	911	889	22	-	97.6	2.4	-
2004(U)	902	865	36	1	95.9	4.0	0.1
2005(V)	879	827	48	4	94.1	5.5	0.5
2006(W)	859	786	69	4	91.5	8.0	0.5
Total(S to W)	4,549	4,365	175	9	96.0	3,8	0.2

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Waves S to W); author's calculations.

Table A1-H

Status of SOEP Households in Sample H

- Households by Year (Wave) and OSM Status -

	Households Total	OSM-Households*)	Mixed Households	NOSM-Households**)	OSM-Households*)	Mixed Households	NOSM-Households**)
Year(Wave)	Number of Households				Shares of OSM Status in percent		
2006(W)	1,506	1,506	-	-	100.0	-	-

*) OSM-Households: households with original sample members only **) NOSM-Households: households with non-original sample members only.

Sources: SOEP (Wave W); author's calculations.