

On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP)

Zur Behandlung von Stichproben-Neuzugängen im Sozio-ökonomischen Panel (SOEP)

*Martin Kroh, Rainer Pischner,
Martin Spieß and Gert G. Wagner*

Abstract

Household panel studies usually provide data on parents, children, and siblings, but not, e.g., on all ex-partners after a divorce or split in a cohabiting unit. This kind of information, however, can be an interesting source for intergenerational research or, more generally, for research on the dynamics of life course. To the best of our knowledge, among the diverse household panel studies, only the German Socio-Economic Panel (SOEP) contains data on individuals entering a sample household in or after wave two also after leaving that household again. This allows all household members to be traced once having lived in the sample household. In this paper, we discuss the rationale for tracing non-original sample members (Non-OSMs) in household panel studies. We also present results on the incidence and thus the relevance of Non-OSMs in the SOEP in general, and for specific research questions in particular.

Zusammenfassung

Haushalts-Panel-Datensätze stellen üblicherweise Daten für Eltern, Kinder und Geschwister zur Verfügung, nicht aber Daten aller Ex-Partner nach einer Scheidung oder nach der Trennung von Lebenspartnern. Diese Information stellt jedoch eine interessante Datenquelle für intergenerationale Untersuchungen beziehungsweise allgemeiner, für Untersuchungen zur Dynamik des Lebenslaufes dar. Unseres Wissens nach sind Daten über Personen, die in Welle zwei oder später in einen Panelhaushalt einziehen und diesen danach wieder verlassen, nur im Sozio-oekonomischen Panel (SOEP) verfügbar. Im SOEP werden alle Personen weiterverfolgt, die einmal in einem Panelhaushalt gelebt haben. Im vorliegenden Papier diskutieren wir die Idee, die den Weiterverfolgungsregeln bezüglich der Neuzugänge nicht-originale Stichprobenmitglieder des SOEP zugrunde liegt. Präsentiert werden weiterhin Häufigkeiten und damit die Relevanz dieser Stichprobenneuzugänge im SOEP im Allgemeinen und für einige Forschungsfragen im Speziellen.

1 Introduction

Prospective household panel studies like PSID in the US, BHPS in the UK, and the German SOEP were originally designed to give a picture of the composition of private households and the well-being of the members of those households at the point in time of the first survey and in the nearer future (e.g., Wagner et al. 1993, 2007). However, the endurance of some panel surveys, such as SOEP with its now 25 years, entails that a large percentage of the first-wave survey participants have now formed new households of their own with other individuals, many of whom were not sampled in the wave one households but joined survey participants' households later, either by birth or by moving in. Furthermore, many of the new individuals who grew into the survey after wave 1 have left these households since. These changes in the composition of households reflect typical dynamics of living arrangements over more than two decades, and it has become more and more apparent that the individual-level data from household panel surveys can be an excellent source for the analysis of these changing relationships.

From a practitioner's point of view, however, these changes prompt the not-so-trivial question of who should be followed up as a sample member of a longitudinal household panel study, and who should not. The main differences between panel studies exist in the tracing rules for adults who join the panel after wave 1. For example, household panel studies like the PSID do not follow individuals who entered after wave 1 and who leave the sampled household again. As an exception to this 'classic' rule, the BHPS follows new sample members if they are the parents of individuals sampled in wave 1. However, if parenthood is deemed to constitute an 'important relation' to an original sample member (short, OSM) in the BHPS (Jenkins/Siedler 2007), thus justifying the tracing of non-original sample members (short, non-OSM), then one must ask why a former marriage or partnership of an OSM to a non-OSM is not sufficiently 'important'.

Whereas all household panel studies 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/Siedler 2007: 6). This is due to the tracing rule adopted for the SOEP, whereby *all* household members aged 17 and older are traced after living in a sample household. This means that even individuals that entered the sample after the corresponding households joined the survey, i.e., non-OSMs, are followed up after they leave the originally sampled household.

Whereas the design of the random first-wave sample and also of later random refreshment samples is usually straightforward just as in any cross-sectional

survey, non-OSMs who join existing households by way of self-selection require adaptation of sampling concepts. These adaptations are reflected in the weighting concepts of a household panel study. In this paper, we report in more detail the tracing rules of SOEP and present results on the incidence and thus the relevance of non-original sample members after 25 years. Section 2 discusses the rationale for tracing all individuals from each sample household instead of original sample members only in household panel studies, and in particular in the SOEP. In Section 3, we briefly discuss the ramifications of the tracing of non-original sample members for the sampling design and thus the weighting scheme of the SOEP. Section 4 concludes.

2 Inclusion and Tracing Rules

In household panel surveys, the primary selected units are households, and the individuals living within these households are additional, implicitly selected units. Because households are 'artificial units', the question of which household and individual, respectively, 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 individuals, the answer becomes much more difficult in the case of household panels. The difficulties arise from the fact that households are defined by a set of individuals, and this set may change over time. For example, let household h be a sample household at time t , consisting of one individual, i , who is a sample member. Suppose that an additional individual, j , moves into this household between t and $t+1$ together with a child. Further assume that individual j leaves household h together with the child between $t+3$ and $t+4$ to form a new household, k , still belonging to the same target population. In our example, individual i would be an original sample member (OSM), i.e., a unit selected and observed in the first wave and, correspondingly, individual j a non-OSM, i.e., a unit not observed in the first wave.

Thus, the set of OSM individuals includes all respondents of wave one, i.e. original participants, and their children living in the observed wave-one-household. Similarly, on an individual level, all household panel surveys consider children born into an OSM household after wave one, as well as children of OSM households living abroad at wave one to be original sample member. Non-OSM units at the household level are all existing not observed households in the population of interest as well as all existing households not in the population of interest. The population of interest, or 'target population', in the case of SOEP are the house-

holds in Germany and the population not of interest are households abroad. Even more general, one may add all not yet existing households to this set. At the individual level this set includes all persons not observed in OSM households, i.e., not-observed individuals existing in the population, those who are existing outside the target population, and even those that were not yet born in wave one outside the sample.

Let us assume that in the above example, t is the first wave. Then, household h is an OSM, but household k is not. If we define only the OSMs as the units to be traced, then only household h is a sample member at times $t = 1, 2, 3, 4$. In addition to the household itself, usually its members are of interest. Turning to the individual level, individual i is an OSM. In wave two, the non-original sample member j enters the household and, depending on the inclusion rule, he or she may or may not be a sample member in wave two and later waves. Usually, individual j would be interviewed in wave two and thus would be a sample member in this wave. However, the difference between the tracing rule adopted in the SOEP and rules adopted for other household panels is that in the SOEP, household k and individual j are sample members even in wave four when they no longer live with the OSM. In the same spirit, the child living in household k becomes a respondent at the age of 17 in the SOEP. The alternative, adopted by other surveys such as PSID and BHPS, would be to ignore household k and individual j (together with the child, who is not a biological child of person i) from wave three on.

The SOEP was started with the 'classic' tracing rule established by PSID that only those households – and all the eligible individuals living therein – are to be surveyed that are inhabited by OSM individuals as well as their children. The SOEP fieldwork followed this classic concept for the first six waves (1984–89). However, it turned out to be difficult for interviewers in the field to distinguish between OSMs and non-OSMs, and a great deal of interview data were deleted after collection. Beginning with wave seven in 1990, the decision was made by the SOEP group in Berlin and the fieldwork organization Infratest Sozialforschung¹ in Munich to survey all eligible individuals living in a sample household, and to follow those living in the sample household at time t from t to $t + 1$. Not eligible are all household members who explicitly refuse to participate and those who did not participate two times in succession. Note that these inclusion and tracing rules imply that even individuals may become sample members that never lived in an OSM household nor ever actually lived with an individual OSM.

1 To the best of our recollection, this was proposed by the head of the fieldwork organization, Bernhard von Rosenblatt.

Table 1 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, including children of those HH (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). Note that Table 1, according to our definition of OSM's in the panel context, treats not yet born children of OSMs and children of OSMs living abroad in wave one as OSMs themselves once they appear in their parents' households. Note also that samples A and B started in 1984, sample C in 1990, sample D in 1994/1995, sample E in 1998, sample F in 2000, sample G in 2002, and sample H in 2006.

Table 1 The Number of Households in SOEP 2006 by Sample and Non/OSM Status

Samples	Number of Households				Proportion of OSM Status in Percent		
	Total	OSM HH ^{*)}	Mixed HH	Non-OSM HH ^{**)}	OSM HH ^{*)}	Mixed HH	Non-OSM HH ^{**)}
A	2,821	1,572	950	299	0.557	0.337	0.106
B	655	392	223	40	0.599	0.340	0.061
C	1,717	1,123	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	3,895	3,394	450	51	0.871	0.116	0.013
G	859	786	69	4	0.915	0.080	0.005
H	1,506	1,506	-	-	1.000	1.000	-
All	12,361	9,490	2,317	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 samples: the most recent sample H from 2006, by definition, includes only originally sampled persons in each interviewed household. In samples 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 samples, A and B, slightly more than half of the households are populated exclusively by respondents who were wave-one household members, children living abroad, and unborn

children of OSM's. Interestingly, more than 10% of the households in sample 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 samples A through G separately.

Obviously, the number and proportion of household inhabited by non-OSM individuals varies over the samples, suggesting that the proportion of households with non-OSM individuals is increasing with the number of panel waves. Exceptions to this pattern are sample B, which is a sample where the household head did not have the German citizenship in 1984, and sample D, a sample of immigrants. Although the results of Table 1 suggest an increasing importance of non-OSMs, one may wonder whether these cases, from an individual perspective, represent a sustained enlargement of the SOEP survey or whether their proven mobility is indicative of a higher attrition rate in the following years. In other words, one may ask whether non-original sample members joining a SOEP household are more volatile than original sample members and thus, whether their contribution is only temporary. 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 2).

Table 2 distinguishes three groups of respondents. The first group is made up of those who were members of the originally sampled households *and* participated in the initial wave of each sample A through G in 1984, 1990, 1994/5, 1998, 2000, and 2002, $t_{0,A-G}$. The second group contains individuals who were members of the originally sampled households in samples A through G but were *not* interviewed in $t_{0,A-G}$. This applies to individuals who were too young to participate in $t_{0,A-G}$, who were not yet born or who were living abroad in wave one but became part of the active sample in one of the following years $t_{>0,A-G}$. The final group contains respondents who were not members of sampled households in $t_{0,A-G}$ and thus participated, like the second group, for the first time in $t_{>0,A-G}$, i.e., a wave subsequent to initial sampling. While the first two groups represent the raw sample of individuals who belong to the original sample of households, the third group covers external entrants to the survey.

Table 2 The Probability of Continued Participation of Persons by Non/OSM-Status

Years After First Interview	Original Sample Members		Non-Original Sample Members
	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	35,899	5,268	6,275
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).

The entries in Table 2 denote 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 samples 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 sixth wave of each individual's initial interview, new sample members have an even somewhat *higher* response probability, with 70%, than interviewees who already lived in the originally sampled households at $t_{0,A-G}$ with 69%. Only in the very long run is the continued participation of first-wave respondents better than in the two other groups. The latter is arguably due to the much lower age of the respondents who enter the SOEP in waves 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 $t_{0,A-G}$ of 45 years). There is, however, no indication that new sample members are dramatically more volatile than those in the original sample of households.

3 The Advantage for Tracing Non-OSMs

In principle, a tracing rule that follows non-original sample members creates a kind of snowball-effect that, for the number of waves going to infinity and a population that is growing at a lower rate than the sample, would lead to the inclusion of all households and individuals living in the target population ('universe'). However, this does not happen at all due to the attrition rates of households and individuals. Much more importantly, the constant growing of non-original sample members into ongoing longitudinal household panel surveys may, in principle, capture some of the dynamics in the underlying population that cannot be reproduced efficiently in panel designs focusing exclusively on first-wave respondents.

As an example, consider the age of individuals. In a longitudinal survey of a fixed set of first-wave respondents ('original participants'), the average age will grow parallel with the length of the panel. Post-stratifying weighting variables would be the only way to compensate the growing difference between the age distribution in the sample and the underlying population. However, in household panel surveys, children of original sample members eventually reach the age of interviewing and thus grow into the sample, and individuals not sampled originally join existing households as well. In principle, these expansions of the sample diminish the differences in the age structure between the sample and the target population.

Table 3 presents the mean age of the subsamples of the SOEP samples A to H: first, it gives the mean age of persons in 2007 who were already interviewed in the first wave of the respective samples. As one would expect, the more distant the year of sampling, the higher the mean age of respondents. Due to age differences between subsamples already existing in the years of sampling, to deaths, and to age-specific attrition rates, the mean age of respondents does not precisely mirror the 'age' of the subsamples. Second, the table shows the mean age of respondents in 2007 who were sampled in the first waves of each subsample but who did not provide an interview. In most cases, these are persons who were 16 or younger in the first wave and grew into the survey in later years. This also includes not yet born children of original sample members. Obviously, the mean age of this group in 2007 is much lower than in the first group. Taking the average of both groups of original sample members shows that the inclusion of these first-wave respondents who did not provide an interview in the first wave significantly reduces the mean age of older samples such as sample A and B of 1984 by about ten years.

Third, the table gives the mean age of respondents in 2007 who did not provide an interview in the first waves of each subsample and who were not sampled at all in wave-one. This group thus consists mostly of individuals who entered

the SOEP sample by moving into an existing household. The mean age of this group in 2007 is between that of the first two groups, and reduces the average age of the cross-section by another two to three years in the oldest samples A and B of 1984 and also C of 1990.

Table 3 The Mean Age of Observed SOEP Respondents in 2007 by Sample and Type of Entrance in the Survey

Sample	Wave	Gross Sample in t_0 and children			Mean Age of Non-OSM's	Total Mean Age
		Mean Age of OSM's with Interview in t_0	Mean Age of OSM's without Interview in t_0	Mean Age of All OSM's. Average of Columns 3 and 4		
A	24	61	28	52	41	49
B	24	57	29	45	38	43
C	18	57	26	49	37	47
D	14	50	22	46	37	45
E	12	55	22	51	38	50
F	8	54	25	51	36	50
G	6	53	24	50	35	49
H	2	51	33	51	34	51
Total		55	26	50	39	49

In the end, the mean age of respondents in all SOEP subsamples in 2007 hovers around 50 years, although samples A and B were drawn more than 20 years prior to the most recent refreshment sample H. The descriptive results in Table 3 thus imply that excluding the non-OSM individuals, i.e., those who enter the sample after wave one, leads to higher mean ages in the sample than in the 'augmented' samples, where these non-OSM individuals are included.

Obviously, the results in Table 3 suggest that ignoring the non-OSMs may lead to selective samples with respect to characteristics of the underlying cross-sectional population, at least with respect to age and variables correlated with age. Ignoring these units would thus require additional modelling of the corresponding missing mechanism at the analysis stage. These corrections usually are performed – as in BHPS and PSID – by way of post-stratified weighting. The inclusion of non-OSMs in principle reduces the variance of cross-sectional weights and thus increases the efficiency of any weighted estimation. Note that for some dynamics in the underlying population, such as new inflows of migrants, it is unlikely that the inclusion of non-original sample members sufficiently captures these developments. This due to the fact that the fusion of households is likely to be a non-

random process, but one that is selective with respect to household characteristics. Since it is rather unlikely that immigrants, who form new households and thus become relevant population units, are affected by this strategy in a rather small number of waves, the German SOEP accounts for immigrants that do not enter existing households through its special 'immigrant sample', sample D (cf. Burkhauser et al. 1997) and through general refreshment samples (cf. Wagner et al. 2007).

Besides the advantage that augmenting the sample with non-original sample members reduces the need for weighting, the broad tracing rules in SOEP facilitate a number of specific research questions that focus on relationships ('dyads'). One prominent example is the intergenerational transmission literature, e.g., the transmission of poverty from parents to children (Jenkins/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/Winkelmann 2005).

Table 4 illustrates that a large portion of the data that permits the matching of individual respondents with their parents, siblings, but also grandparents, uncles, and aunts involve non-original sample members. For instance, more than half of the individuals with information on their parents were not interviewed in wave one, most because they were too young (16 years or younger). Restricting the units of analysis to first-wave respondents clearly reduces the number and the types of relationships.

Table 4 Shares of Respondents with Interviews on Relatives in the SOEP and Total Figures on Respondents by the Type Entrance in the Survey

	Original Sample Member		Non-Original Sample Members	Total	
	Person with Interview in t_0	Person without Interview in t_0		All	N
Parents	0.42	0.52	0.06	1.00	11,366
Siblings	0.34	0.60	0.06	1.00	8,394
Grandparents	0.25	0.59	0.16	1.00	729
Uncles/Aunts	0.02	0.75	0.23	1.00	238
Total	0.75	0.13	0.12	1.00	48,016

Note. Reading example: Of the 11,366 SOEP respondents interviewed having either a father or mother who was also interviewed, 42% were already interviewed in wave one of the respective SOEP sample, 52% were part of the wave one gross sample but did not participate in wave-one, and 6% entered the gross sample after wave one. *Data Source.* SOEP data distribution 2008.

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 economic respectively social circumstances ('environment') on human behaviour and well-being. Schimmack/Lucas (2007), for example, used SOEP data to analyze the well-being of the same couples during marriage and after divorce.

Table 5 reports the incidence of partnership dissolutions in the SOEP for which data exists on both partners before *and* after the separation. About half of these cases would not have been possible to observe if the tracing rules had covered first-wave respondents only. SOEP's tracing rules allow partnership dissolutions to be observed more than once in several individuals: without following up both original and non-original sample members, particularly the consequences of such repeated separations would be impossible to investigate.

Table 5 Dissolution of Partnerships with Interviews with both Partners before and after Separation by Type of Entrance into the Survey

	Both Persons	Mixed Couples	Both Persons	Total	
	Original Sample Members		Non-Original Sample Members	All	N
1 st Dissolution for Both Partners	0.47	0.45	0.06	1.00	838
2 nd Dissolution for at least one Partner	0.43	0.49	0.08	1.00	357
3 rd + Dissolution for at least one Partner	0.19	0.68	0.13	1.00	69

As yet another example, consider the living conditions, income, and well-being of widows/widowers. Every widow/widower in the sample, whether an OSM individual or not, can contribute to this kind of research. Over the course of time in a long-running panel, the percentage of non-OSM widows/widowers in the sample increases because more and more widows/widowers enter the sample through marriage to OSM individuals. Ignoring these non-OSM individuals would lead to a loss of a great amount of information, for example, about the 'final resting place' of the deceased OSM individuals (cf. Gerstorff et al. 2008), a topic that has received increasing attention in recent years.

4 Implications of Inclusion Rules for Weighting

The choice of inclusion and tracing rules not only is important for facilitating substantive research, but clearly also has ramifications for the weighting scheme. Weights as usually delivered with archived data sets, are generated to allow users of the data sets to compensate for different selection and response probabilities. At the heart of most theoretically justified weighting approaches is the probability of observing the corresponding unit at the corresponding point in time (e.g. Robins/Rotnitzky/Zhao 1995; Wooldridge 2002). Selection probabilities are usually known from the sampling design, whereas response probabilities have to be estimated. Obviously, units are only observed if they are selected into the sample and if they participate. The observation probabilities can be modelled by a sequence of known or estimated conditional probabilities. In principle, this is the procedure adopted for the SOEP (Kroh/Spieß 2008; Goebel et al. 2008: 88–96). It allows a straightforward incorporation of Non-OSM in the weighting scheme.

If the rule is adopted to trace all eligible individuals inhabiting a sample household from wave t to wave $t+1$ and to survey all households in $t+1$ inhabited by such a former-wave sample individual during that period of time, i.e., if we want to also include non-original sample members once they form a new household, the task is to derive, model, and estimate the probabilities of observing households that were observed in t and $t+1$ and, for new households, of observing households that were not observed in t but are observed in $t+1$. Modelling and estimating the observation probabilities of households with move-ins from t to $t+1$ can be demanding, depending on whether the new household member was (a) part of the sample, (b) part of the population of interest, i.e., living in Germany, but not in the sample, and (c) not part of the population in t , i.e., living abroad.

However, deriving the observation probabilities for split-off households is, under assumptions, generally straightforward. Without additional nonresponse on the individual level, the probability of observing individuals from a sample household is equal to the probability of observing that particular household (Galler 1987).

One requirement for estimating observation probabilities for a variety of groups is that the number of cases in the sample be sufficient to allow the estimation of group-specific effects on response behaviour. The smaller the number of observations within relevant groups with different response behaviours becomes over time, the more one has to rely on the models used to generate the weights. Furthermore, ignoring cases in $t+1$ that are in the sample at time t introduces a missing data problem. If these missing cases (households or individuals) are not missing completely at random (e.g., Rubin 1987), then the missing data process cannot gen-

erally be ignored. Thus, the inclusion respectively the tracing rules adopted in the SOEP to include and trace all eligible household members prevents modelling and estimation of this aspect of the (non-)response models, which in turn are part of the process of deriving or estimating observation probabilities. Again, the efficiency of the weighting scheme benefits from the broad inclusion rules.

This point is illustrated in Table 3, where inclusion of those individuals that entered the sample later than wave one lowered the mean age of most of the samples to the level of the latest sample, sample H, selected in 2006. At least part of a possible selectivity with respect to age and variables correlated with age can be overcome by allowing these units to grow into the sample. Thus, it is not necessary to model this aspect of the (non-)response process, relying on (untested) assumptions.

5 Conclusions

The German Socio-Economic Panel Study (SOEP) is a household panel survey with a different tracing rule than other household panel studies like PSID and BHPS. PSID drops Non-OSMs once the OSM leaves a sample household. BHPS does so as well, but with the exception of those Non-OSMs who are parents of OSM-children. Only SOEP traces all Non-OSMs. That is, even if a new, i.e., non-original household member leaves the sample household, the SOEP considers this individual an established part of the survey and continues to trace her subsequent living arrangements. This strategy, originally adopted on a non-theoretical basis to make fieldwork more efficient, turned out to enrich the data significantly because it reduces the necessity to weight and opens up new research questions.

If the classical tracing rule is applied, the analyst has not only to compensate (if necessary) for differing sampling probabilities, first-wave nonresponse and attrition in later waves to derive valid inferences, but also for ignoring non-OSMs. Given the results reported in Section 2, these missing units must be assumed to be not missing completely at random, which requires appropriate weighting. A simple strategy, however, is to trace and include these units into the sample.

Non-OSMs will eventually make up a large portion of the respondents and households in any long-running panel design. Analyses reported in this paper do not suggest that these cases are more volatile in their participatory behaviour than OSMs. Furthermore, 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/Lucas 2007). In addition, tracing Non-OSMs is helpful for the analysis of the terminal phase of life (cf. Gerstorff et al. 2008). The value of this tracing rule will increase further after the introduction of 'exit interviews' into the SOEP dealing with the terminal phase of life of respondents who have passed away (Kröger 2008).

6 References

- Burkhauser, R. V., P. Giles, D. R. Lillard and J. Schwarze, 2005: Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. *Journals of Gerontology Series B- Psychological Sciences and Social Sciences* 60 (5): 238–246.
- Burkhauser, R. V., M. Kreyenfeld and G. 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, J., M. Francesconi and Thomas Siedler, 2006: Intergenerational Mobility and Marital Sorting. *The Economic Journal* 116: 659–679.
- Galler, H. P., 1987: Zur Längsschnittgewichtung des Sozio-ökonomischen Panels. S. 295–317 in: H.-J. Krupp and U. Hanefeld (eds.): *Lebenslagen im Wandel*. Campus: Frankfurt and New York.
- Gerstorff, D., N. Ram, R. Estabrook, J. Schupp, G. G. Wagner and U. Lindenberger, 2008: Life Satisfaction Shows Terminal Decline in Old Age. Longitudinal Evidence from the German Socioeconomic Panel Study. *Developmental Psychology* (in press).
- Goebel, J., M. Grabka, P. Krause, M. Kroh, R. Pischner, I. Sieber and M. Spieß, 2008: Mikrodaten, Gewichtung und Datenstruktur der Längsschnittstudie Sozio-ökonomisches Panel (SOEP). *Vierteljahrshefte zur Wirtschaftsforschung* 77: 77–109.
- Haisken-DeNew, J. P. and J. R. Frick, 2005: DTC Desktop Companion to the German Socio-Economic Panel (SOEP). Version 8.0 - Updated to Wave 21. (Berlin: DIW). <http://www.diw.de/documents/dokumentenarchiv/17/38951/dtc.354256.pdf> (7.11.2008)
- Jenkins, S. P. and T. Siedler, 2007: Using household panel data to understand the intergenerational transmission of poverty. DIW Discussion Paper 694.
- Kröger, K., 2008. "Exit-Interviews" in SOEP und SHARE. DIW Data Documentation (forthcoming).
- Kroh, M. and M. Spieß, 2008: Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP 1984 until 2007). DIW Data Documentation 39.
- Lucas, R. A., 2007: Adaption and the Set-Point Model of Subjective Well-Being. *Current Directions in Psychological Science* 16 (2): 75–79.
- Rendtel, U., 1995: *Lebenslagen im Wandel: Panelfälle und Panelrepräsentativität*. Frankfurt/New York: Campus.
- Robins, J. M., A. Rotnitzky and L. 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.
- Rubin, D. B., 1987: *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Särndal, C.-E., B. Swensson and J. Wretman, 1992: *Model Assisted Survey Sampling*. New York: Springer.
- Schimmack, U. and R. Lucas, 2007: Marriage Matters: Spousal Similarity in Life Satisfaction. *Schmollers Jahrbuch* 127 (1): 105–111.

- Schwarze, J. and R. Winkelmann, 2005: What can happiness research tell us about altruism? Evidence from the German Socio-Economic Panel. DIW Discussion Paper 475.
- Wagner, G. G., R. V. Burkhauser and F. 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., J. R. Frick, and J. Schupp, 2007: The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. Schmollers Jahrbuch 127 (1), 139-169.
- Wooldridge, J. 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)	1,056	1,056	-	-	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)	7,232	6,574	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.

Contact Address:

Martin Kroh
 Deutsches Institut für Wirtschaftsforschung
 Abteilung SOEP
 Mohrenstraße 58
 D-10117 Berlin
 mkroh@diw.de