

Attrition analysis of income data

Working Paper 2_09

Swiss Household Panel, FORS

February 2009

by

Ursina Kuhn

Correspondence to:
Panel Suisse de Ménages, FORS, c/o Université de Lausanne, Bâtiment Vidy, CH-
1015 Lausanne

e-mail: ursina.kuhn@fors.unil.ch

Introduction

This paper is an extension on Attrition analysis of the Swiss Household-Panel (SHP) by Voorpostel (2009). Departing from the same framework it looks specifically at attrition on income variables.

Attrition effects on income variables merit special attention for various reasons. Firstly income measures have shown to be affected by panel attrition in previous studies (Socio Economic Panel SOEP, European Community Household Panel ECHP). Bias hasn't however been assessed for the SHP so far. Secondly, information on income is an important for data analysis and especially for inference on the population. Thirdly, better understanding of attrition effects is one basis for the decision on a new future sample of the SHP. One option (among many others) would be to oversample certain income groups or to draw a special sample on a particular income group.

This paper extends the attrition analysis by Voorpostel (2009) in the SHP, by looking at attrition effects for income variables. In her paper, Voorpostel looks at two types of variables. The first group consists of demographic variables (gender, age, education, Swiss nationality, region, urbanization, civil status, children present in household, home owner). The second group of variables consists of attitudes and behavior regarding social involvement (take part in clubs or other groups' activities, participation in federal polls, general trust in people, interest in politics, influence on government policy, satisfaction with health).

As a typical socio-economic variable, income is part of neither of these groups. Like most socio-economic variables it is a constructed variable relying on a series of conceptualizations and definitions. Compared to socio-demographic variables, there are many difficulties with practical implications to be considered. Information on socio-demographic variables are relatively easily available on different levels. Direct questions on demographic variables in individual questionnaires are relatively unproblematic so that non-response is relatively rare and responses are relatively reliable. Additionally, information at demographic variables is available in the SHP even for individuals who did not take part of the individual questionnaire, but where the household-structure questionnaire (grid) has been filled out. Furthermore, for many demographic variables the population distribution is known, so that even consequences of bias from initial non-response to the survey can be assessed. Income variables are more problematic in all these respects. Item-non response to income questions is quite high¹ and there is a considerable amount of imprecision and measurement error for various reasons. Additionally, for non respondents little is known on their income, as there is no information available from the household grid. Also regarding the distribution within the Swiss population we know very little so far. There are other difficulties with income variables because questions in data collection have changed over time for the SHP I sample having started in 1999. For the SHP II sample, method of data collection has not changed since its imposition.

In this paper, we refer to Voorpostel (2009) for the review of the literature, choice of framework of analysis and the construction of the attrition variable. While we will perform mostly the same analyses, some adaptations are necessary to apply it to income variables. The reason for this lies mostly in the continuous measurement level of income data. We will thus

¹ Bias from item non-response is relatively well researched in the literature. With sophisticated imputation, its bias can be reduced considerably. On unit non-response in income variables there is in contrast much less literature available.

firstly present some graphical assessment of attrition variables for the overall income distribution. Then we will look at attrition effects for both income quintiles (because they are not influenced by outliers) as well as income inequality measures.² We then will also assess attrition with regard to income quintiles and variables of social integration. Before these analyses we will –briefly – address the question of sampling special income groups.

Sampling particular income groups

Apart from a general assessment of attrition in income variables, the aim of this paper is to provide a basis for the discussion on whether a future sample of the SHP could or should be drawn with respect to certain income groups. It does however not outlay and assess advantages and disadvantages of these different procedures. Nevertheless, I will consider here two independent points on attrition effects, namely substantive research interest and sampling with respect to income variables.

Research interest

Drawing a subsample of a specific income group has enormous implications for the possibility of data analysis. The previous samples of the SHP and most other scientific surveys are random samples. This implies that the sample size for specific sub groups is usually too small, even though the overall sample size of the SHP is relatively high for Switzerland. Specific samples of subpopulations would however allow analyzing that particular subgroup.

With respect to persons with low income, this would be an interesting option with potentially great use among the research network. “Poverty, living conditions, quality of life” is an important research domain in the SHP research network and there have been various inquiries of potential data user with regard to sample sizes for low income groups (recipients of invalidity pensions, recipients of social assistance, single mothers) which could be realized with such subsamples. This “demand” needs however further assessment.

For sub sampling upon income in general, it is important to consider also the positioning of the SHP with respect to SILC-Switzerland. SILC, having been constructed on the basis of the SHP-questionnaire and with the same sampling frame, overlaps considerably with the SHP. However, income variables are central in the SILC-survey and one of the central purposes of SILC is calculate indicators on poverty. Giving more importance to income variables in the SHP, would therefore also mean to increase similarity of the SHP and the SILC survey. However, sub-sampling of certain income-groups (or any other group) would allow for further distinguishing SHP and SILC data, because this would provide the basis for different research questions.

Sampling

There are two different possibilities how to oversample certain income groups. The first is to apply an appropriate sampling frame and the second one is screening.

² In literature, the effects of attrition and item-non response on income mobility are also often looked at. This could be done in a further version of this paper.

Theoretically, sampling could be done on the basis of the tax register as a sampling frame, where selection of cases is based on taxable income. For reasons of data protection, this strategy is however unrealistic. Another obstacle lies in the current method of data collection, as both addresses as phone-numbers have to be available to conduct a CATI survey (and to send a letter beforehand).

Also the approach of case selection by screening might be very problematic. Amount of income still being a delicate question in Switzerland and item-non-response being rather high (in comparison to other survey questions), we have to expect the sample to be already biased highly in first wave, as we are likely to “miss” many potential cases.

Measuring attrition and income

In the SHP, many different variables on income are available. At the household level, there are variables for total household income and for equivalent household income. At the individual level, apart from total personal income, there are between 4 and 13 different income sources (depending on the year). The number of persons touching a particular income source varies considerably.

We will restrict the following analysis to three income variables: net equivalent household income, net total personal income and net working income (resp. employment income for SHP II). These variables have been chosen as they are the ones most frequently used by data users and because either a sampling frame or screening would have to be applied to any of those variables for practical reasons.

The various attrition groups have been constructed by Voorpostel (2009) and contains the following categories:

- “Always in”: Respondents who completed an individual interview in every wave
- “Ever out”: Respondents who did not complete all waves, but who were present at least once and this was at least at one of the two latest waves
- “Lost”: Respondents who did not respond in the last two waves
- “Deceased”
- “Out”: Respondents that were known to be institutionalized or had left the country

For the other variables, answers were taken from the first wave in which the respondent participated. There is however an exception for the income variables in the SHP I. Due to reasons of data quality, responses given in 2001 or 2002 were preferred over those in 1999 and 2000 if they were available.³

The sample used consists of all individuals who responded to the individual questionnaire in at least one of the waves.

Findings in literature

In a study of attrition-bias for income Behr, Bellgardt and Rendtel (2003) analyze data in the European Community Household Panel (ECHP). They find that Response rate is increasing with income in the northern countries, especially in the UK, Denmark and France and sharply declining with income in southern countries, especially in Italy, Portugal and Greece. They

³ The reason is that in 1999 and 2000, income plausibilisation has not been done exactly within the same framework/program as in later waves. Additionally, in 1999 income from old-age pension has not been asked in the questionnaire, which could influence overall income of respondents.

also estimate bias on poverty, income inequality and mobility under panel attrition. *Poverty rates* are underestimated in Northern countries as a selective attrition (in DE, UK, DK, IR, NL). The *poverty gap* is underestimated in 7 countries, but only in Germany and Greece significantly. As regards the *Gini-coefficient*, there is no clear tendency of over- or underestimation of inequality when using only the overall respondents. In five countries the attrition results in a significant underestimation, while in Spain and Greece inequality is significantly overestimated. For *income mobility* there is no clear tendency of over- or underestimation of mobility due to attrition. In the UK and ES mobility is underestimated, in PT and IT mobility is overestimated.⁴ Overall, the authors come to the following conclusion: “overall we find that the effects of attrition are rather mild. Nevertheless for some countries significant biases were found for different measures. The conjuncture stated by Rendtel (1995) that attrition is linked to mobility cannot be confirmed by our results. The effect goes in either direction for different countries. ... We find that attrition effects to be very heterogeneous across the different national subsamples of the ECHP.”

An analysis on the basis of the Finnish sample of the ECHP is in line with those results. Sisto (2003) finds “a non-response bias towards underestimation of mobility in household income and measures of inequality”. But she also finds that “in most cases the initial wave non response causes more differences than the attrition and the size of the attrition bias appears to be of moderate size. There is no apparent trend towards larger bias due to attrition during the panel”.

For other household panels, there is a large body of literature for bias of item-non response and how this bias can be corrected through imputation. No references for the effect of attrition on income and income distribution have been found so far.⁵ In attrition analysis of the SOEP (Kroh and Spiess 2007: 20) however, there is one table on attrition of all first-wave respondents by income quintiles (household income). It shows that attrition is decreasing with increasing income, but overall differences are rather small.

Attrition on income distribution

First of all, we look at the overall income distribution, which allows inspecting on attrition effects over the whole income distribution. However, diagrams are restricted to positive income up to 150'000 CHF. Due to outliers, diagrams would hardly be readable otherwise. Also for greater readability, we restrict the presentation to the contrast of the groups “always in” and “lost” and do not consider the intermediate category “ever out”.

Figures 1-6 indicate, that for both SHP I and SHP II, persons with lower income more frequently attrite from the SHP than respondents with higher income. For personal income and working income, the distribution of the “always in” respondents is slightly flatter than the one of the “lost” group. The overall shape of the two groups is however rather similar for both groups. Overall, we can say that some attrition effects are visible, but that they do not considerably distort income distribution.

⁴ In a very similar article by the same authors, similar results are presented. The conclusion with no clear trends towards income mobility remains (Behr, Bellgardt, Rendtel (2006).

⁵ Research efforts in literature have however not been extensive so far. This can be done later.

SHP I

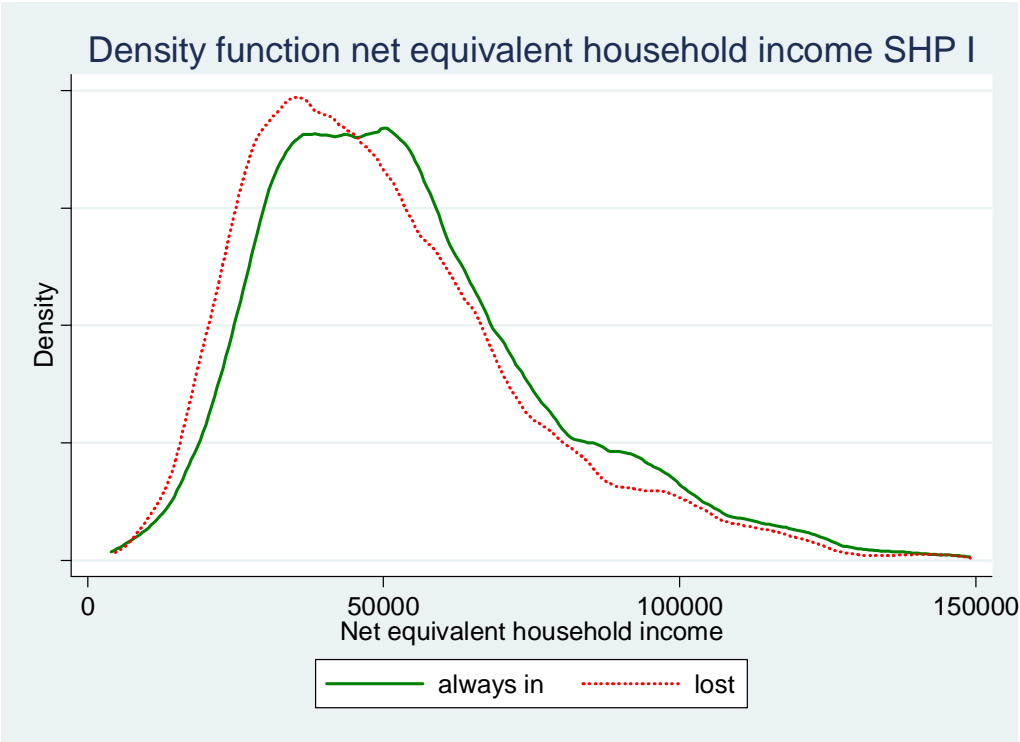


Figure 1 : Density function yearly net equivalent household income, SHP I

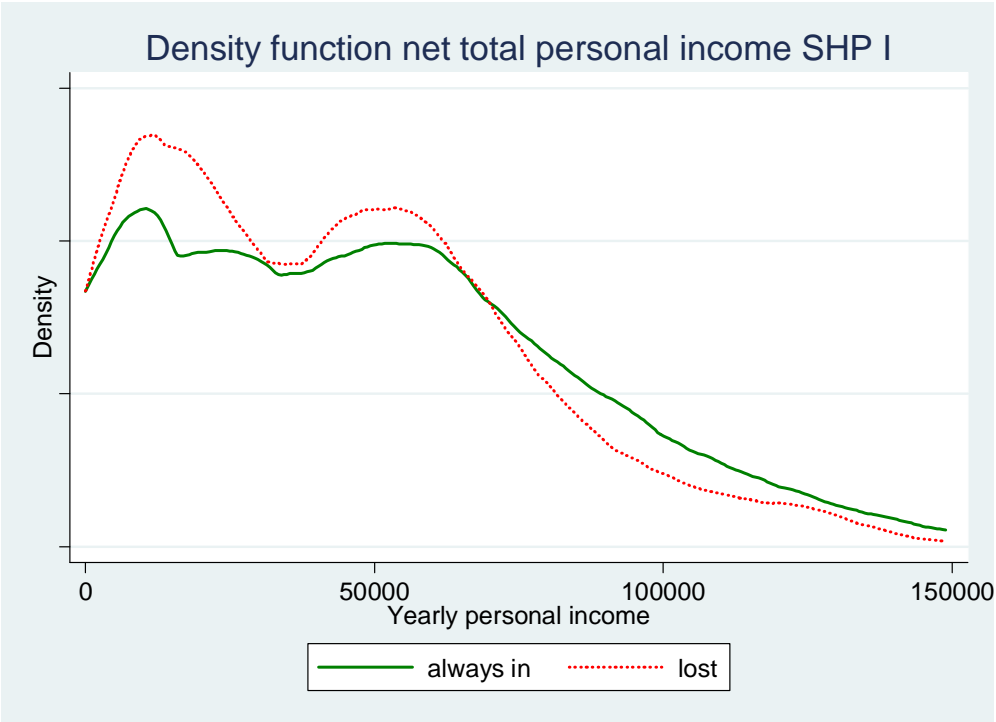


Figure 2 : Density function yearly net total personal income, SHP I

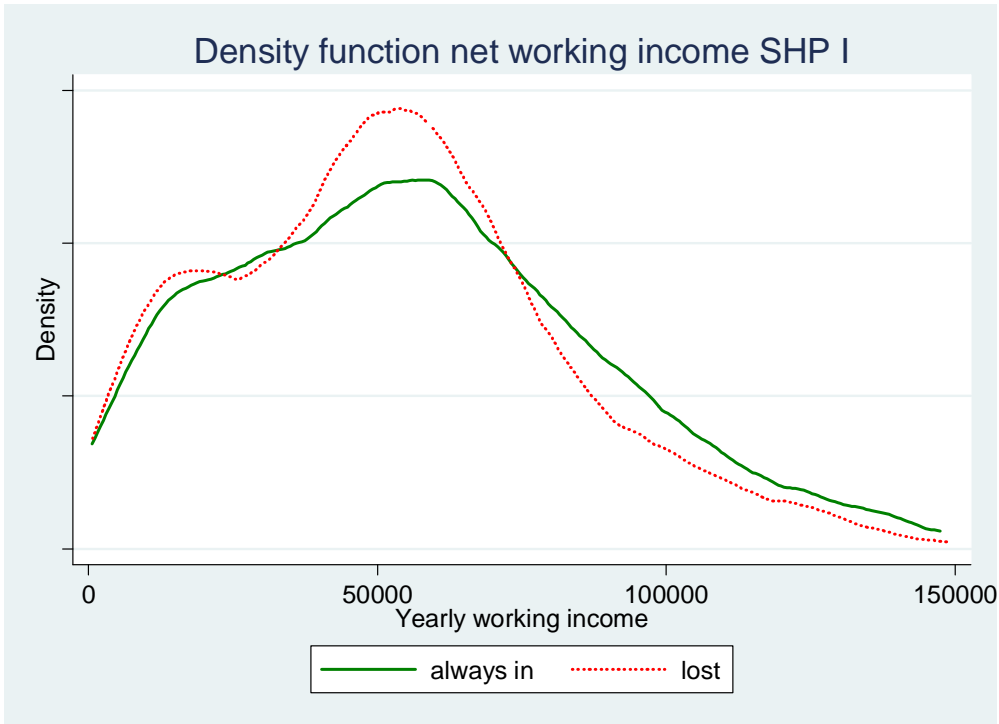


Figure 3 : Density function yearly net working income, SHP I

SHP II

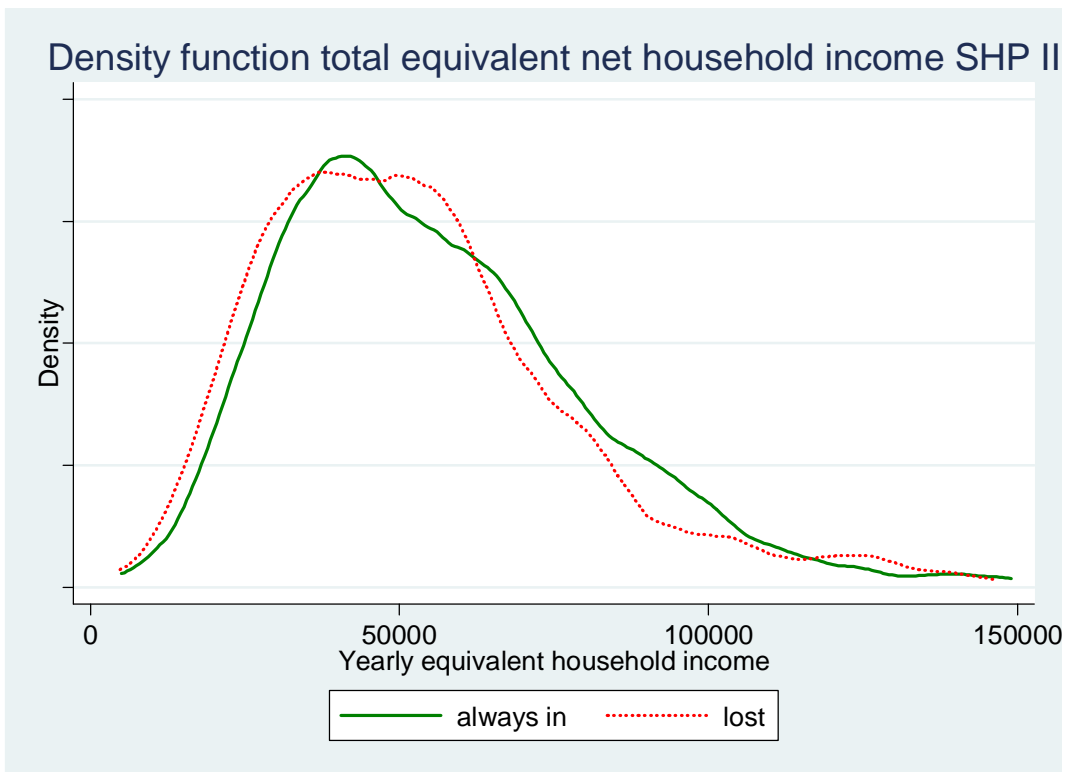


Figure 4 : Density function yearly net equivalent household income, SHP II

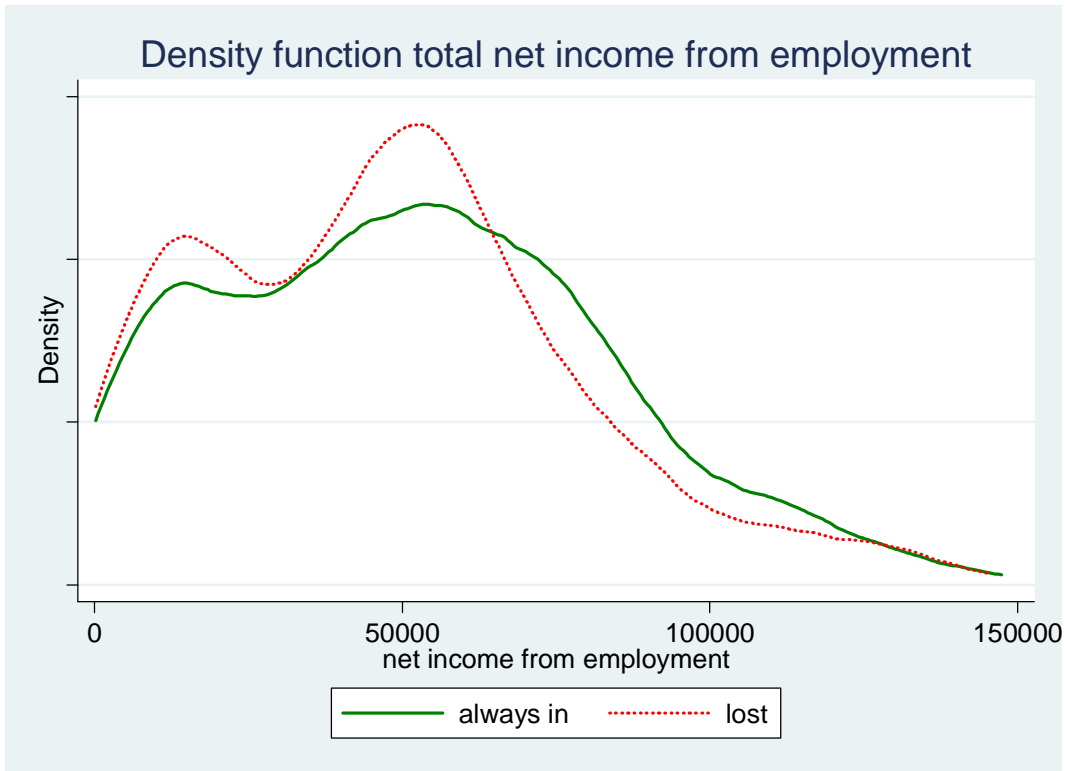


Figure 5 : Density function yearly net total personal income, SHP II

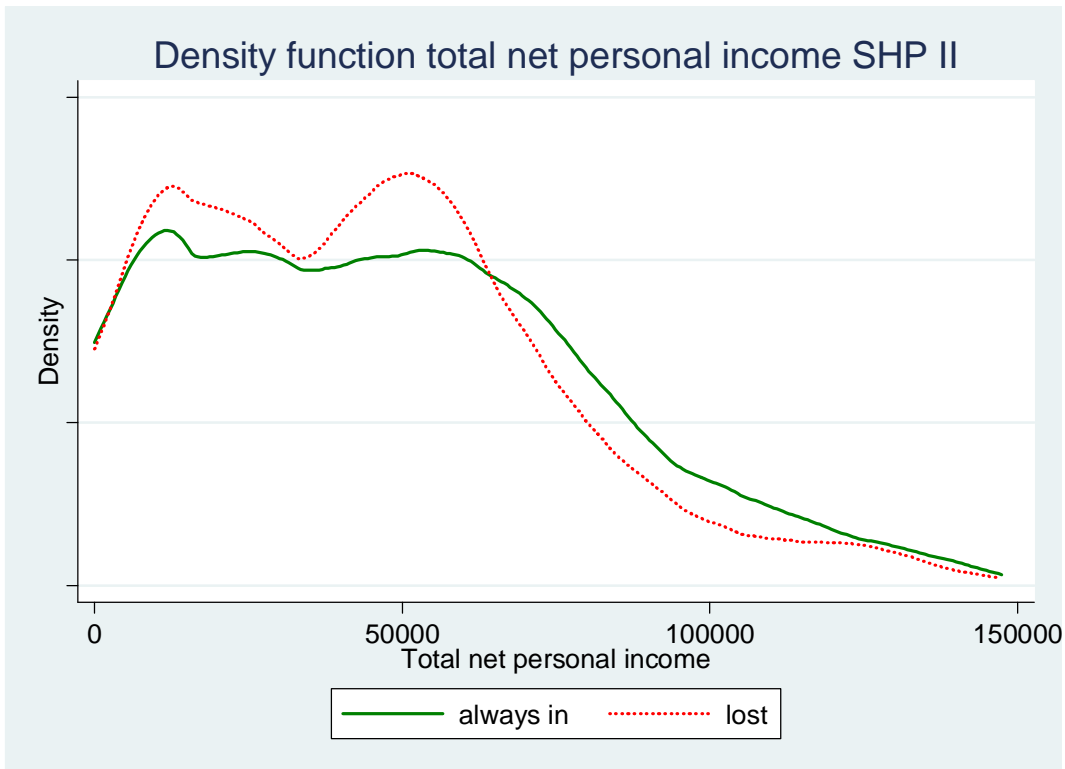


Figure 6 : Density function yearly net working income, SHP II

Attrition on income quintiles

Generally, attrition is highest in the lowest or in the second lowest quintile for the income sources considered (see tables 1 and 2). This is in line with findings of the SOEP (Kroh and Spiess 2008). The only exception is working income in SHP I where person in the third quintile are most likely to drop out from the panel. However this does not hold for the SHP II and also here, attrition is lowest for in the highest quintile as in all other cases. Usually, the group “everout” takes an intermediate position between the “always in” and the “lost” group. As the Cramer’s V indicates, the relationship between the attrition variable and income quintiles is quite small.

SHP I

	always in	everout	lost	deceased	out of sample	Total
equivalent household income in quintiles						
1. Quintile	15.2	20.9	22.0	32.4	20.8	20.0
2. Quintile	18.7	20.9	20.4	11.3	26.4	20.0
3. Quintile	21.6	18.8	19.6	31.0	15.1	20.0
4. Quintile	21.8	19.7	19.4	15.5	9.4	20.0
5. Quintile	22.8	19.7	18.7	9.9	28.3	20.0
n	2166	1908	3952	71	53	8150
<i>Cramer's V</i>		0.09	0.09			
total personal income						
1	18.3	22.7	19.6	11.1	43.8	20.0
2	18.7	19.7	22.9	33.3	28.1	20.8
3	18.9	19.5	20.7	15.6	12.5	19.7
4	20.4	18.3	19.8	13.3	15.6	19.6
5	23.8	19.8	17.0	26.7	0.0	20.0
n	2271	1755	2672	45	32	6775
<i>Cramer's V</i>		0.07	0.09			
working income						
1. Quintile	18.7	20.3	21.0	33.3	30.8	20.1
2. Quintile	20.4	19.7	20.4	11.1	7.7	20.2
3. Quintile	18.4	19.6	22.8	11.1	30.8	20.5
4. Quintile	19.1	19.5	19.4	11.1	30.8	19.3
5. Quintile	2.4	20.9	16.5	33.3	0.0	20.0
n	1506	1110	1725	9	13	4363
<i>Cramer's V</i>		0.04	0.10			

Table 1 : SHP I Response groups by income quintiles

SHP II

	always in	everout	lost	deceased	out of sample	Total
total net equivalent household income						
1. Quintile	13.87	21	27.46	25	16.67	20.15
2. Quintile	20.69	21	17.78	25	25	19.85
3. Quintile	20.63	20.91	19.67	16.67	33.33	20.42
4. Quintile	21.75	17.98	18.23	8.33	16.67	19.58
5. Quintile	23.06	19.11	16.87	25	8.33	20
n	1687	1057	1322	12	12	4090
<i>Cramer's V</i>		0.10	0.17			
total personal income						
1. Quintile	15.8	22.5	23.6	8.3	18.2	20.0
2. Quintile	18.8	21.9	19.9	41.7	36.4	20.1
3. Quintile	20.9	18.1	20.3	8.3	18.2	19.9
4. Quintile	20.2	19.3	20.3	25.0	27.3	20.1
5. Quintile	24.3	18.1	15.9	16.7	0.0	19.9
n	1611	1004	1254	12	11	3892
<i>Cramer's V</i>		0.11	0.13			
income from employment (age from 18 to 65)						
1. Quintile	15.8	20.6	25.0	0.0	28.6	20.0
2. Quintile	20.1	20.9	19.2	50.0	28.6	20.0
3. Quintile	19.8	19.7	21.1	0.0	14.3	20.2
4. Quintile	20.7	20.6	18.2	0.0	28.6	19.9
5. Quintile	23.6	18.3	16.5	50.0	0.0	19.9
n	1032	623	812	2	7	2476
<i>Cramer's V</i>		0.08	0.13			

Table 2 : SHP II Response groups by income quintiles

Attrition on income inequality

For income, it is not only relevant which income level is most affected by attrition. It is also important to assess how distributional characteristics, such as income inequality is affected by attrition.

However, here we cannot apply the same procedure as before in contrasting the different response groups. It doesn't make sense to compare the Gini-Index of respondents (always in) and for drop-outs, as such a comparison would not allow inferring on attrition bias. It is e.g. possible that the drop outs are more equal than the respondents, but that nevertheless, the "always in" group could be more equal than the whole population. We therefore calculate inequality measures for the group "always in" (1), for "always in" plus "everout" (2) and for "always in" plus "everout" plus "lost" (3). Comparing results for respondents in comparison to the whole population, allows giving a rough indication on the bias created by attrition.

We will calculate three different inequality measure for these groups: the Gini index, the ratio of the highest decile relative to the lowest decile (p90/p10) and the ratio of the highest quartile relative to the lowest quartile (p75/p25). Table 3 and 4 indicate that attrition tends to decrease inequality measures. The effect is however very small and might not be significant (\$\$ are there significance test on inequality measures?). Generally, this bias is higher for measures that are more sensitive to extreme values (as the Gini-Index or the p90/p10 ratio).

SHP I

	always in	always in & everout	always in & everout & lost
<i>equivalent household income in quintiles</i>			
Gini	0.266	0.275	0.28453
p90/p10	3.264	3.469	3.475
p75/p25	1.843	1.881	1.923
<i>net working income</i>			
Gini	0.355	0.370	0.365
p90/p10	7.313	7.823	7.879
p75/p25	2.469	2.468	2.404
<i>net total personal income</i>			
Gini	0.444	0.464	0.462
p90/p10	57.618	57.222	33.854
p75/p25	3.941	3.972	3.800

Table 3 : SHP I measures of income inequality for different response groups

SHP II

	always in	always in & everout	always in & everout & lost
<i>equivalent net household income</i>			
Gini	0.28	0.29	0.29
p90/p10	3.34	3.47	3.48
p75/p25	1.90	1.92	1.92
<i>net income from employment</i>			
Gini	0.36	0.36	0.37
p90/p10	9.75	9.78	9.55
p75/p25	2.68	2.80	2.82
<i>net total personal income</i>			
Gini	0.42	0.44	0.44
p90/p10	20.83	24.74	21.36
p75/p25	3.64	3.68	3.59

Table 4 : SHP II measures of income inequality for different response groups

Attrition on income quintiles and social integration

Looking at attrition effects on variables on social integration, we find variables on social integration influence attrition. This holds for both participation in clubs or groups and political interest, for nearly all income quintiles and all income variables considered in the SHP I. In the SHP II effects are not as strong, but no clear pattern with respect to income quintiles can be seen.

SHP I

Participation in clubs or groups	Always in	Ever out	Lost
<i>equivalent household income</i>			
1. Quintile	56.4	46.6 **	43.5 ***
2. Quintile	58.9	53.5	43.4 ***
3. Quintile	62.5	55.2 *	48.7 ***
4. Quintile	61.2	56.4	49.9 ***
5. Quintile	58.6	50.5 **	45.3 ***
<i>total personal income</i>			
1. Quintile	67.0	54.8 ***	58.0 **
2. Quintile	60.1	50.3 **	42.5 ***
3. Quintile	55.6	43.3 ***	37.6 ***
4. Quintile	53.7	59.8	50.3
5. Quintile	65.6	56.3 **	55.6 **
WORKING INCOME			
1. Quintile	66.0	51.6 **	48.9 ***
2. Quintile	58.0	48.9 *	37.6 ***
3. Quintile	56.3	49.1	44.7 **
4. Quintile	57.8	59.7	53.3
5. Quintile	68.0	58.2 *	58.1 **

Remarks: *p<.05 **p<.01 ***p<.001

Table 5 : SHP I Proportion of respondents indicating participation in clubs or groups by income quintiles for different response groups

Mean interest in politics (0–10)	Always in	Ever out	Lost
EQUIVALENT HOUSEHOLD INCOME			
1. Quintile	4.9	4.3 *	4.1 ***
2. Quintile	5.2	4.7 *	4.4 ***
3. Quintile	5.5	5.2	4.7 ***
4. Quintile	5.6	5.2	4.8 ***
5. Quintile	5.9	6.1	5.5 *
TOTAL PERSONAL INCOME			
1. Quintile	4.4	4.4	4.3
2. Quintile	5.5	4.9 **	4.7 ***
3. Quintile	5.2	4.6 **	4.5 ***
4. Quintile	5.7	5.6	5.0 ***
5. Quintile	6.7	6.5	6.2 **
WORKING INCOME			
1. Quintile	5.4	5.1	4.9 *
2. Quintile	5.1	4.5 *	4.6 *
3. Quintile	5.1	4.9	4.4 ***
4. Quintile	5.9	5.6	5.2 ***
5. Quintile	6.7	6.9	6.3 *

Table 6 : SHP I Mean score on interest in politics (0-10) by income quintiles for different response groups

Remarks: *p<.05 **p<.01 ***p<.001

SHP II

Participation in clubs or groups			
	Always in	Ever out	Lost
<i>Net equivalent household income</i>			
1. Quintile	44.0	40.5	44.5
2. Quintile	49.6	45.1	40.9 *
3. Quintile	49.7	52.0	48.1
4. Quintile	59.4	44.7 **	47.7 **
5. Quintile	54.5	55.0	43.5 **
<i>Net total personal income</i>			
1. Quintile	58.3	52.7	48.0 *
2. Quintile	46.4	42.7	42.4
3. Quintile	45.1	40.7	35.4 *
4. Quintile	48.8	45.9	49.8
5. Quintile	61.5	58.2	53.8
<i>Net income from employment</i>			
1. Quintile	54.6	45.3	43.1 *
2. Quintile	43.5	48.5	41.0
3. Quintile	41.7	39.8	43.9
4. Quintile	54.7	47.7	52.7
5. Quintile	60.7	56.1	50.8

Table 7 : SHP II Proportion of respondents indicating participation in clubs or groups by income quintiles for different response groups

Remarks: *p<.05 **p<.01 ***p<.001

Mean score on interest in politics (0–10)	Always in	Ever out	Lost
<i>equivalent household income</i>			
1. Quintile	5.2	4.9	4.3 ***
2. Quintile	4.9	4.6	4.8
3. Quintile	5.6	5.4	4.8 **
4. Quintile	6.1	5.4 **	5.1 ***
5. Quintile	6.5	6.4	6.0
<i>total personal income</i>			
1. Quintile	4.8	4.8	4.2 *
2. Quintile	5.5	4.8 **	4.7 **
3. Quintile	5.4	5.3	4.9
4. Quintile	5.9	5.3 *	5.1 ***
5. Quintile	6.8	6.7	6.6
<i>working income</i>			
1. Quintile	5.4	4.9	4.7 *
2. Quintile	5.3	4.8	4.5 *
3. Quintile	4.7	4.9	4.2 *
4. Quintile	6.0	5.6	5.4 *
5. Quintile	6.9	6.6	6.8

Table 8 : SHP II Mean score on interest in politics (0-10) by income quintiles for different response groups

Remarks: *p<.05 **p<.01 ***p<.001

Conclusions

Respondents with lower income are generally underrepresented in the panel due to attrition effects. This results in an overestimation of average and media income, but also in an underestimation of income inequality. These results are in line with findings for northern countries in the ECHP and for the SOEP. However, even if there is a bias on estimates on income variables, this bias is quite small and does not effect the overall distribution of income variables to a great extent.

As regards attrition of effects of social integration we draw the same conclusions as Voorpostel. The bias does not vary much across the different income quintiles and different income variables considered. Oversampling or sub-sampling of lower income groups does thus not resolve the problem of attrition on variables on social integration. Subsampling on income variables should does not be mainly motivated by attrition, but there should also be a substantive interest on particular subsamples.

There are various ways, in which this paper could be enhanced. Most of all, it would be important to link item non-response and attrition as there are several indication that they are related. It would e.g. be important to know whether tem-non response and unit-non response point into the same direction. Also in terms of imputation for unit non-response there is not much that has been done so far. Another aspect that so far has been neglected is to test to what extent weighting helps to reduce bias from attrition with regard to income variables. This would be crucial to assess the possibility to use SHP data also for cross-sectional descriptive statistics such as income inequality. How are we able to disentangle time-trends and attrition effects? Furthermore, an the effect of attrition on income inequality has not been looked at so far and should be done so in the future.

References

Behr, Andreas, Egon Bellgardt and Ulrich Redtel (2003). Comparing poverty, income inequality and mobility. Chintex working paper 12. J.W. Goethe-Universität Frankfurt/Main.

Sisto, Johanna (2003). Attrition Effects on the Design Based Estimates of Disposable Household Income. Chintex working papter 9. Statistics Finland.

Kroh, Martin and Spiess, Martin (2008). Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2007). DIW Berlin Data Documentation 39.

Voorpostel, Marieke (2009). Attrition in the Swiss Household Panel by demographic characteristics and levels of social involvement. SHP-Working Paper.