Session Number: Poster Session 1

Time: Monday, August 25, 17:30-18:30

Paper Prepared for the30th General Conference of

The International Association for Research in Income and Wealth

Portoroz, Slovenia, August 24-30, 2008

Representative wealth data for Germany: The impact of methodological decisions around multiple imputation and the choice of the aggregation unit

Joachim R. Frick, Markus M. Grabka, Eva M. Sierminska

For additional information please contact:

Name	Joachim R. Frick
Affiliation	SOEP at DIW Berlin
Full mailing address	Mohrenstr. 58, 10117 Berlin, Germany
Email address	jfrick@diw.de

This paper is posted on the following website: http://www.iariw.org

Representative wealth data for Germany:
The impact of methodological decisions around multiple imputation and the choice of the aggregation unit

by

Joachim R. Frick*

Markus M. Grabka**

Eva M. Sierminska***

* Dr. Joachim R. Frick, German Socio-Economic Panel Study (SOEP), DIW Berlin, TU Berlin, IZA Bonn; DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany, jfrick@diw.de, Tel.: +49-30-89789-279, Fax: +49-30-89789-109

** Dr. Markus M. Grabka, German Socio-Economic Panel Study (SOEP), DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany, mgrabka@diw.de, Tel.: +49-30-89789-339, Fax: +49-30-89789-109

*** Eva M. Sierminska, Ph.D., CEPS/INSTEAD, BP 48, L-4501 Differdange, Luxembourg, eva.sierminska@ceps.lu, Tel. +352 585 855 643, Fax. +352 585 560,

Abstract:

The definition and operationalization of wealth information in population surveys and the corresponding microdata requires a wide range of more or less normative assumptions. However, the decisions made in both the pre- and post-data-collection stage may interfere considerably with the substantive research question. Looking at wealth data from the German SOEP, this paper focuses on the impact of collecting information at the individual rather than household level, and on "imputation and editing" as a means of dealing with measurement error.

First, we assess how the choice of unit of aggregation or unit of analysis affects wealth distribution and inequality analysis. Obviously, when measured in "per capita household" terms, wealth is less unequally distributed than at the individual level. This is the result of significant redistribution within households, and also provides evidence of a significant persisting gender wealth gap.

Secondly, we find multiple imputation to be an effective means of coping with selective non-response. There is a significant impact of imputation on the share of wealth holders (increasing on average by 15%) and also on aggregate wealth (plus 30%). However, with respect to inequality, the results are ambiguous. Looking at the major outcome variable for the whole population—*net* worth—the Gini coefficient decreases, whereas a top-sensitive measure doubles. The non-random selectivity built into the missing process and the consideration of this selectivity in the imputation process clearly contribute to this finding.

Obviously, the treatment of measurement errors after data collection, especially with respect to the imputation of missing values, affects cross-national comparability and thus may require some cross-national harmonization of the imputation strategies applied to the various national datasets.

Keywords: Wealth, Item non-response, multiple imputation, SOEP

JEL classification: D31, C81, I32

Contents

List of Tables

List of Figures

1 Motivation

The definition and operationalization of wealth information in population surveys and the corresponding microdata requires a wide range of more or less normative assumptions. However, such decisions may interfere considerably with the substantive research question at hand, as has been shown, for example, by the research on imputation of missing income data and its effects on income inequality and mobility (see e.g., Frick & Grabka 2003, Biewen 2001). Also the choice of the time interval chosen to measure income may cause income inequality results to differ between annual and monthly income data (see Wagner et al 2001 for the case of Germany using SOEP data, and Böheim and Jenkins 2000 for results based on the BHPS in the UK). Looking at wealth data from the German SOEP as prepared for the LWS database, this paper focuses on two of the aforementioned issues: the impact of imputation and editing, and the choice of the aggregation unit.

Firstly, missing data due to item non-response is a major problem for all micro-data, and an even more acute problem for economic outcome variables such as income and wealth. Not only does the decision to use imputation rather than weighting play a role (or assuming missing at random by not dealing with missing data), but even more so, the choice of the imputation procedure. Using micro panel data on annual income from the UK BHPS, the Australian HILDA survey, and the German SOEP, Frick & Grabka 2006 provide empirical evidence that incorporating longitudinal information significantly improves the quality of the imputation results. These surveys make use of single imputation techniques (for income variables), which may be criticized for not sufficiently considering the uncertainty embedded in the imputation process. This phenomenon is taken into account by Kennickell (1998), who uses data from the US Survey of Consumer Finances (SCF) to show the relevance of applying multiple imputation techniques, as well as by Spiess and Goebel (2005) based on a unique linked dataset from Finnish survey and register data.

Secondly, in contrast to many other surveys, the SOEP surveys wealth data at the individual level (as does the BHPS), making it possible to investigate how the choice of the aggregation unit affects wealth distribution and inequality analysis (including the decomposition of inequality). The main hypothesis here is that household-based information is more equally distributed given the internal redistribution process embedded in the “pooling and equal sharing assumption” implicitly made in such welfare-oriented analyses. Following from this, a major advantage of individualized wealth data is the opportunity to investigate the size and determi-

nants of a possible “gender wealth gap”. If results differ for the two approaches (based on aggregated or household-level versus individual-level information), this may provide useful information for other countries where no individualized wealth data is available, controlling for similar household structures (size and composition).

Section 2 of this paper briefly describes the underlying SOEP data in general, as well as how the wealth data was collected in the survey year 2002, including the editing and imputation techniques applied to this data. Section 3 of the paper deals with inequality analysis, focusing on the choice of the aggregation unit, comparing results on inequality derived from both aggregation levels. Section 4 deals with the scope and impact of multiple imputations. In the final section, we conclude, giving specific consideration to the relevance of our results for cross-national comparative research.

2 The data

The Socio-Economic Panel Study (SOEP) is a representative longitudinal survey of individuals living in private households in Germany (Frick & Haisken-DeNew 2005). Recent developments and plans for upcoming waves are described in Wagner et al (2006). The survey was started in West Germany in 1984 and extended to East Germany in 1990. The initial sample included over 12,000 respondents, with all individuals aged 17 and over in each household being interviewed. In the years 1995, 1998, 2000, 2002 and 2006, new sub-samples were drawn which approximately doubled the initial sample size. The sample analyzed here comprises approximately 12,700 households with about 24,000 respondents (plus their children) surveyed in the year 2002. That year, the individual questionnaire included a special module focusing on wealth (see Appendix A1).

This section included questions on seven different wealth components:

- owner-occupied property (including debt)
- other property (including debt)
- financial assets
- private pensions (including life insurance and building savings contracts)
- business assets
- tangible assets
- consumer credit

Potential shortcomings of this fairly extensive wealth questionnaire arise from the exclusion of cars in the measure of tangible assets[1] and the lack of information about pension entitlements through both company pensions and the German statutory social pension fund (*Gesetzliche Rentenversicherung*). This information is difficult to collect because individuals may not be aware of the amount they might be entitled to through these forms of social security. A further restriction comes with the use of a lower threshold of 2,500 euros for financial and tangible assets and consumer credits. This was set in order to reduce the burden on respondents by not asking them to state negligible amounts. As such, the overall measure of total wealth and even more so the share of wealth holders are likely to be understated and biased against very small wealth holdings. A further limitation of the SOEP 2002 wealth questionnaire arises due to the lack of wealth data on children. The SOEP only surveys individuals aged 17 and over. Thus any wealth held by younger persons is not considered here, although it may have been captured if a household-based questionnaire was used. In any case, given the minor relevance of wealth holdings by children, this aspect can be ignored.

Despite these shortcomings, when the total wealth of private households measured by SOEP 2002 is compared with the national balance sheets, the survey does quite well in several categories (see Appendix 2). Housing wealth components match very well, as do net business assets. The biggest discrepancy is in the more heterogeneous categories of financial assets and tangible assets, but in these two cases the questions ask respondents only to report balances over 2,500 euros and do not refer to all the components found in the balance sheets. The overall value of the "financial assets" owned by a given person might consist of numerous single items and forgetting one of them yields a higher probability of understating the true value. This is less likely in the case of housing wealth, given that most people do not own more than one home.

As in other surveys, the SOEP has also encountered relatively low response rates with wealth questions. However, there are two types of non-response which are important in this context. The first occurs when for an otherwise responding person a subset of information (i.e., a single variable) is missing. This is referred to as item non-response (INR). As an example, the share of item non-response for the market value of financial assets is about 16%. The second type is the partial unit non-response (PUNR) and refers to cases where one or more, but not

[1] The SOEP asked for tangible assets in the form of gold, jewelry, coins or valuable collections but not for the value of all of an individual's personal belongings, as usually considered for wealth in national balance sheets.

all persons in the household refuse to answer the questionnaire. [2] Obviously, when aggregating information across all adult household members, such partial unit non-response behavior most likely yields an understatement of the true aggregate, if it is not corrected by means of imputation.

As such, there is a need to find out about any selectivity of measurement error (and thus a need for editing and imputation). We do so by estimating the probability for "inconsistency or INR". However, in order to analyze such issues, we must control for potential selectivity into the state of ownership (of assets and debts). For example, more highly educated individuals are also more likely to hold assets (or debts). Such selection can be controlled for here by means of a Heckman selection correction (see Heckman 1979). We run two models estimating the probability of measurement error in a comprehensive measure of "total assets" and of "total debts" (in each case controlling for sample selection). All results are presented in Appendix 3.

Basic findings for the selection model include:

- Assets (and debts) are more common among persons of higher age (although decreasing among the elderly), males, persons with high education, and those living in rural areas. Controlling for employment status, we find the self-employed and civil servants to be more likely to possess assets, while the unemployed, pensioners and individuals out of the labor force are less likely to do so.
- From a methodological point of view, we find clear indications of interview mode effects: all respondents who deviate from the standard PAPI (paper-and-pencil interview) method show a higher probability for ownership of assets or debt. The number of interviews does not correlate with ownership of assets and does have a marginally negative effect on debt.

With respect to the correlates of the probability for inconsistency / item non-response:

- Male respondents are less likely to show such measurement error.
- Higher education reduces the probability of any type of measurement error; low education is also found to increase the risk of inconsistent answers.
- For civil servants we find the expected negative effect, indicating the increased probability for "complete and true" information.

---

2 There is a third type of non-response: unit non-response, which refers to missing information for the whole observation unit "household". In panel surveys, this phenomenon is often dealt with by means of weighting

- Self-employed persons appear to have more problems providing complete information on the value of their assets; however, this is not true for debts.
- Again, variables focusing on the interview situation significantly contribute to variance explanation: self-administered interviews (as well as CAPI interviews) yield a higher need for imputation and editing and the number of interviews increases the probability of "complete and correct" information on assets, indicating a positive learning effect of repeated interviewing and increased confidence in the interviewer.

Before the data can be imputed for item non-response, it needs to undergo extensive checking for consistency and plausibility. If values are not found to be plausible or consistent they must be edited. In the following we use the term "editing" to refer to cases in which an observed non-missing value is changed into another value: for example, when a couple co-owns a given wealth component, such as owner-occupied property. Here, the SOEP questionnaire asks for (an estimate of) the current market value of the house/flat as well as the percentage share owned by the individual. As such, the market value mentioned by both partners referring to the very same object should coincide by and large. Furthermore, if both partners are the sole owners of this property, their respective shares should add up to 100%. Any deviation along these lines is to be considered measurement error and may be corrected for through some type of "editing", in contrast to "imputation", which is carried out in the case of missing information due to item non-response.

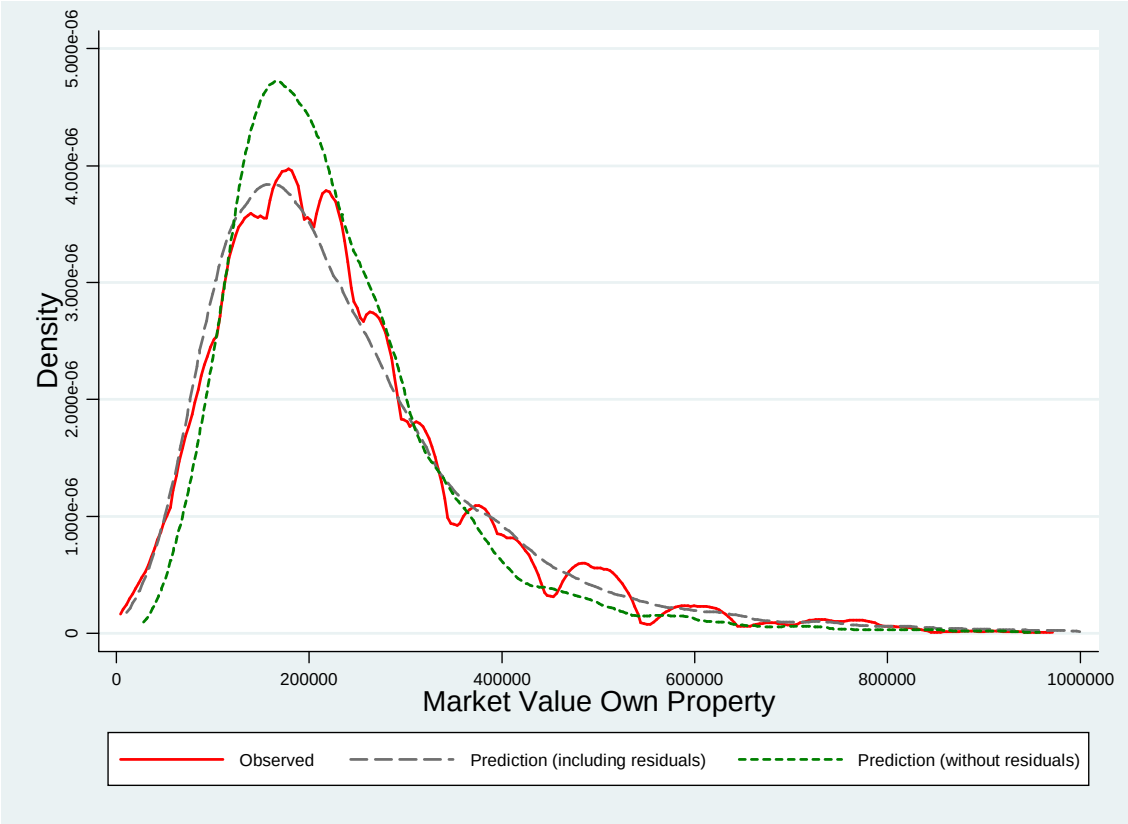
The principal approach for imputing missing market values of the various wealth components in the SOEP is the use of a maximum-likelihood based Heckman selection model controlling for sample selection. In the context analyzed here, sample selection may occur if a respondent refuses to give a valid metric value conditional on his or her answer to a filter question about holding a specific wealth component. These regressions also adequately control for possible regional clustering effects. Finally, in order to better incorporate the uncertainty of the imputation process, a randomly chosen error term is assigned to the regression-based prediction (=imputation). Repeating this process five times yields a multiply imputed data set with five implicates.

In what follows, we demonstrate this principle in the imputation of missing values for the market value of owner-occupied property. Based on the population with valid observed information on the market value of their owner-occupied property, Figure 1 compares kernel

following attrition analysis, where the weighing factor is derived from the estimation of the probability of dropping out (that is, the inverse of the drop-out probability).

density plots for the originally surveyed information and the respective predicted values in order to validate the quality of the imputation procedure. The green long-dashed curve gives the density of the imputed values with, and the grey short-dashed curve without randomly chosen residuals in contrast to the observed market values (red solid curve). Obviously, neglecting residuals yields a distinct regression-to-the-mean phenomenon, that is, a strong underestimation of the variance. However, the comparison of the true distribution with that of the prediction including the error terms for the very same observations shows a high degree of coincidence[3], also providing confidence in the quality of out-of-sample predictions for those with missing values.

Figure 1:
Predicted market values for owner-occupied property and the application of randomly chosen residuals[4]



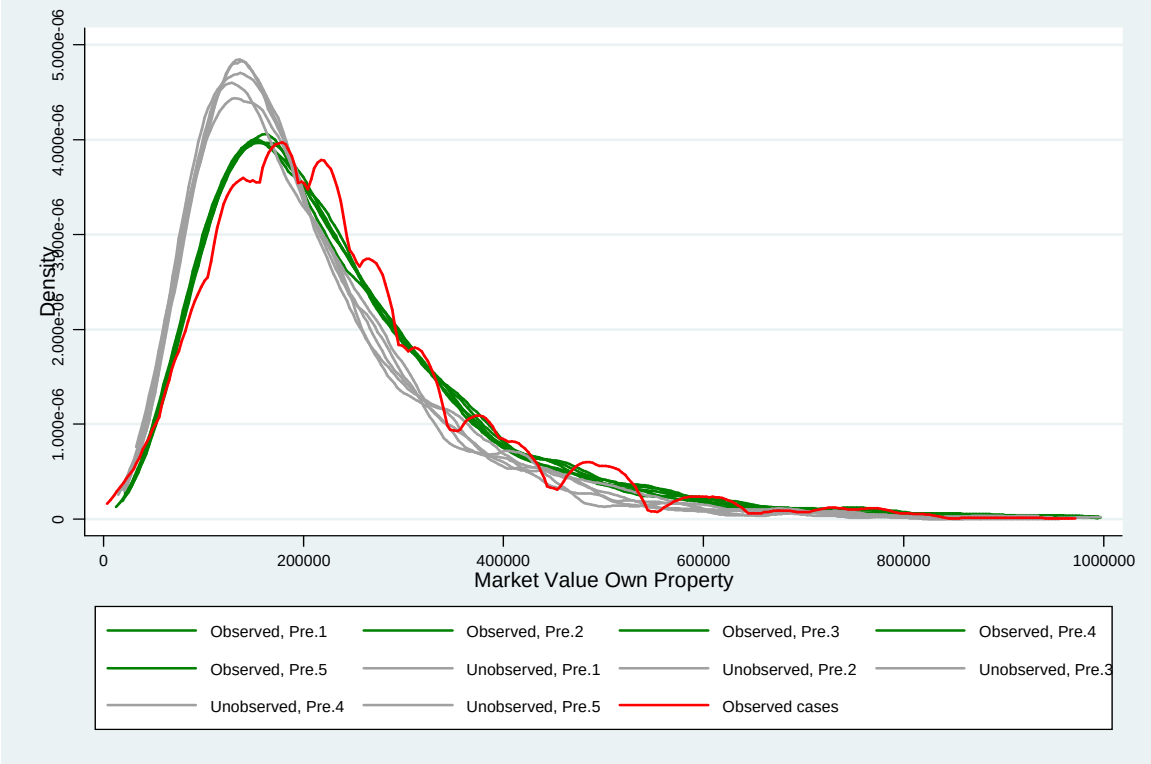
Again, based on the market values of owner-occupied property, Figure 2 demonstrates significant differences between the distribution of the truly observed cases and those which needed to be imputed due to item non-response. Firstly, the five kernel density estimates for the five multiple imputations in cases with INR (these "unobserved predictions" are given as

[3] The distribution of the predicted values consistently lies inside the two-sigma confidence band of the true distribution (not shown in figure for the sake of readability).

[4] Values higher than one million euros are trimmed in this figure. Displayed are the values of all households with an observed market value (5,104 households).

solid grey lines) vary much more than those for the observed cases (green lines). More importantly, the two sets of distributions are distinctively different from each other: the distributions of the five implicates for the imputations are more compressed and are shifted to the left as well. This indicates that the cases with INR are not missing at random (as already shown in the section above) and most likely more prominent in the lower part of the wealth distribution. In fact, these lower market values originate from homes situated in older buildings and in more rural areas, since these are smaller in size and more often occupied by elderly persons with long tenure.

Figure 2:
Comparing multiply imputed cases with item-non-response (unobserved predictions) and fictitious implicates for observed cases (observed predictions) in contrast to the observed cases[5]



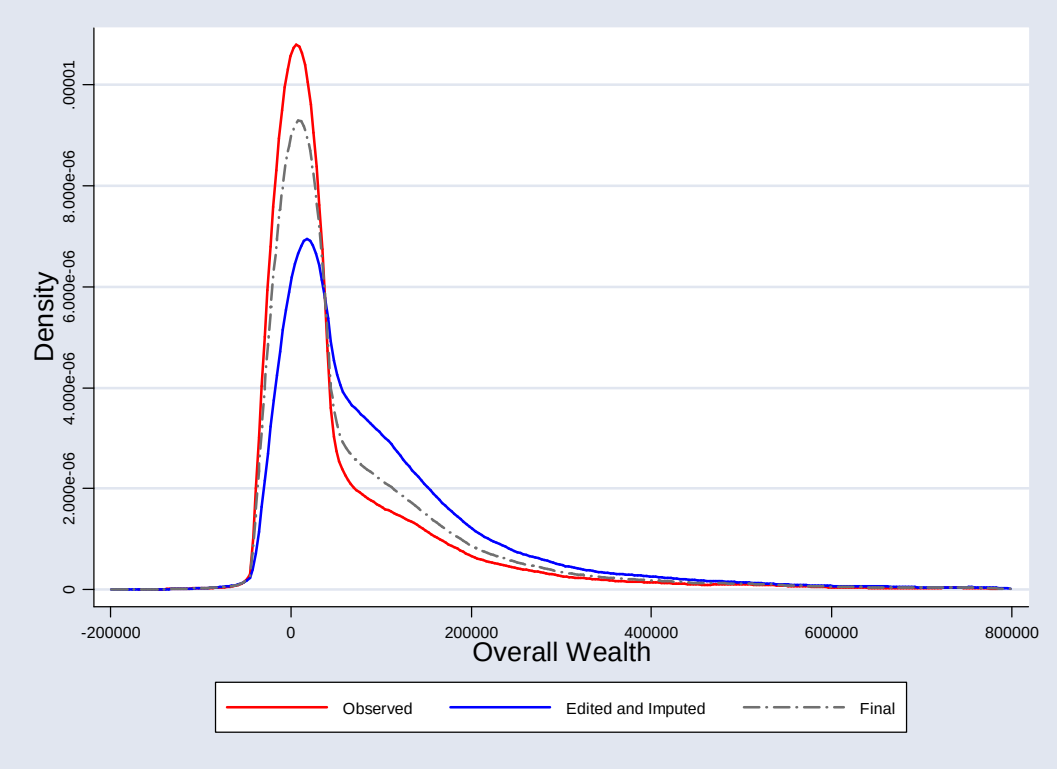
Further details on the editing and imputation strategies are given in Frick, Grabka and Marcus (2007); Appendix 4 gives a general overview of the scope of editing and imputation for the various wealth components.

Figure 3 shows the distribution of *imputed* and *observed* values for overall (net) wealth, which is an aggregate measure of assets and liabilities, as well as the resulting *final* distribution considering all observations. We find distinct differences, particularly close to the zero

[5] Values higher than 1 million euros are trimmed in this figure. Implicates incorporate randomly chosen residuals.

mass as well as in the range of 50,000 to 150,000 euros, that is, the distribution of the imputed cases is shifted to the right and so is the final distribution.[6].

Figure 3.
Comparing observed and imputed values for overall wealth



It should be noted that imputation of missing wealth data in the SOEP is performed at the individual level. However, this individual information can be aggregated at the household level for further analysis. In the following two sections, we will focus on the effect of both (dis)aggregation, as well as editing and imputation, on wealth inequality in Germany in 2002.

3 The “editing and imputation” process

This section deals with scope and impact of multiple imputations on wealth components in the SOEP 2002. First, we examine the incidence of item non-response as a source of editing and imputation, and determine the population share affected by wealth components. Then we focus on the relevance of editing and imputation for aggregate wealth. Finally, we turn to the impact of imputations on summary statistics and inequality measures. For the sake of robustness, we use a wide range of distribution statistics and several inequality measures.

[6] Values less than -200 thousand and more than 800 thousand euros are trimmed in this figure. The distributions of the five implicates as well as for the resulting composite overall wealth variable are more or less identical.

In our analysis, we need to distinguish between total population, the population holding the relevant wealth component, and the observed population that provides complete non-missing information. Our analysis in this section is based on individual-level data and unless otherwise noted, refers to the individual’s personal share in ownership of his or her main residence, investment real estate, financial assets, mortgages, and other mortgaged properties. It is also important to note that some wealth components are only collected for balances of 2,500 euros or more (total financial assets, tangible assets, and non-housing debt).

In Table 1 we find that the effect of imputation and editing for the whole population and total values (not shares) varies from a low 2.3% for other property, followed by business assets, tangible assets, owner-occupied property, and financial assets, up to almost 20% for private pensions. In the aggregate, the imputations and editing compounds to over 30% for total assets. The variation in the imputation rates indicates that individuals’ willingness to respond to asset questions varies with the type of asset. In the liabilities category, the effect of imputation is 50% lower—about 15% in the aggregate. About 6% of the population requires editing or imputing other property debt, followed by other debt and around 12% for the main property debt. The highest prevalence of editing occurs for owner-occupied property and main property debt. In Part B of the table, we focus on the population with positive wealth component holdings. There we find that a greater share of the population is subject to imputation and editing, although to differing degrees than before, suggesting that most imputed and edited values are positive (this is not ex-ante clear given that the imputation process also includes the imputation of filter questions, thus allowing the imputation routine to generate a value of zero).

The impact of imputation[7] on the share of population holding wealth components (asset participation) is found in Table 2. Asset participation increases by almost 50% for business assets and tangibles due to imputation (but participation is low in general; additionally, tangibles are only recorded for balances over 2,500 euros). For most other assets and liabilities the effect is in the range of 16-24% and only 3% for private pensions.

Whereas the population share of imputed or edited wealth data varies across components, we find that on average, about 25-30 % of wealth has been imputed across wealth components (Table 3). The exception is business assets, where nearly 50% of wealth is either imputed or edited. This seems to be the component people are most reluctant to report, or more likely, the one they find most difficult to estimate. For debt, the average figure is 27%, with 31% for

[7] From this point on, the term “imputation” will refer to both “editing and imputation.”

main property, 25% for other property debt, and 18 % for other debt (consumer credits) being imputed or edited.

We find that the share of wealth imputed across components does not vary to a large extent. Figure 4 indicates that across the wealth distribution, the imputed wealth in each decile as a share of wealth in that decile does not vary to a great extent either. For the most part, it is around 40-50%, with 20-30% for the lower deciles. When we look at total imputed wealth, about 60% belongs to the top decile. Less than 5% of total wealth is imputed below the seventh decile: 7% in the 7^{th}, 12% in the 8^{th}, and 19% in the 9^{th} decile.

Summary statistics are presented in Table 4. The median for the whole population is not affected by imputation, and in most cases imputation and editing changed the median by less than 10% among those holding the particular wealth component. This is not at all surprising given that ownership of several components is less than 50% and that most of the imputations are for the richest 10% while the rest are spread out more or less evenly across the distribution (Figure 4). One exception are tangible assets, among which we observe an 18.6% change in the median. Imputation significantly affects the overall population means of wealth components in a positive direction, in effect doubling it for business assets and increasing it by around half for most of the other components. The change in the conditional mean is much smaller, in the range of 1-8%, less significant, and negative, except for business assets and other property debt.

The effect of imputation on selected percentiles of wealth is presented in Table 5. Given that for most assets and debt we do not observe holdings until the median of the distribution, the greatest effect of imputation is for the higher percentile groups. Except for main property and main property debt, editing is performed for only a few observations. Apart from tangible assets and non-housing debt, there is a small effect of imputation at the top of the distribution.

For the whole population, imputation has a significant effect reducing inequality as measured by the Gini and the half-squared coefficient of variation (HSCV) (Table 6). The HSCV is always considerably more reactive than Gini. This can be explained by the imputation process, where values are added in the upper tail of the distribution, increasing the number of observations and effectively reducing inequality in the upper end of the distribution. We also redo this exercise on the basis of observations with positive values only, and find that although inequality is reduced, the imputation process now has an ambiguous effect on inequality. For owner-occupied property (and the associated debt), financial assets and pension assets the effect of imputation has a smaller effect on inequality, whereas it has a larger effect on

investment property, business assets, and other property debt. Once again the effect in most cases is not significant.

Up to now we have focused mainly on components of net worth and the effect of imputation on their summary statistics. Next we turn to overall wealth or net worth, which is a summary measure created as the sum of all assets less liabilities. Table 7 provides summary information on observed and final net worth. As was the case for the components of wealth, mean net worth increases by 35.7% due to the imputation process and by 12.8% for those with positive net worth. Across the distribution, as before, percentiles close to the median are affected most, although we also see an effect at the bottom of the distribution due to imputation for the fifth percentile. The change in inequality is ambiguous. The Gini is reduced and the half-squared coefficient of variation more than doubles. The overall proportion of the poor measured by the Foster, Greer and Thorbecke measure FGT (for alpha=0, that is, the head count ratio) is reduced by 5%. The average normalized poverty gap (FGT1) and the average squared normalized poverty gap (FGT2) are each reduced by about 38 %.

4 The “choice of aggregation unit”

If provided with the choice, many researchers will choose individual-level wealth information over the household level. Wealth data at the individual level is a better match to other welfare-oriented data on income or satisfaction, for example, and if available for multiple waves, individual wealth data vastly improves longitudinal and mobility analysis, especially in the case of household split-offs. Unfortunately, the bulk of the available survey data provides only household or family-level data but not individual data. The SOEP 2002 offers a unique opportunity to compare the individual and the household levels.

As mentioned before, the “equal sharing assumption” implicit in household-based welfare analysis is thought to give rise to a more equal wealth distribution than individual-level analysis. In this section, we explore this hypothesis using the German SOEP data collected at the individual level and then aggregated to the household level. First, we look at the effect of the choice of the aggregation unit on a wide range of (simple) distribution statistics and then decompose inequality by age and gender. We also examine the sensitivity of inequality measures to outliers at the top of the distribution. We distinguish three types of units of analysis: individual, household per capita based on individual data, and household per capita based on household-level data. The household per capita approach based on individual data uses

households that completed interviews,[8] whereas the household per capita approach based on household data may include individuals that did not complete interviews (PUNR). As a result, the household size and wealth aggregates may differ slightly across these groups, being larger in the latter.

Table 8 reveals that data analyzed at the individual level (comparing (a) and (b)) yields greater wealth for the top 10% of the distribution. The minimum increases and the maximum decreases when using the household per capita approach and the means and sum are approximately constant.

Comparing (a) and (b) also shows the expected decrease in inequality (especially for top-sensitive measures as HSCV) and poverty due to the redistribution process within private households. The Gini coefficients are significantly different from each other, whereas the half-squared coefficients of variation are not.

Comparing the two household per capita approaches (comparing (b) and (c)) we find there is an increase in inequality for the "pure" household approach (c). Once again, only the Gini coefficient is significantly different across the different unit types.

As a check, we also examine (in the right panel of Table 8) the corresponding information for disposable income using the exact same definition of population as used in columns (b) and (c) (due to the lack of completely individualized income, we need to exclude children from the "pure" individual approach). We find that, as in the case of wealth, inequality increases when we change the unit of observation from individual to household per capita information. Similarly, we observe an increase in net worth for upper quantiles: although the wealth mean and median falls, the opposite occurs for income. This is partly caused by the fact that there are virtually no households with zero income, but many with zero (or negative) net wealth.

The "pure" individual approach (left panel of Table 8, column (a)) appears to be the most appropriate way to perform this type of welfare-based analysis. However, for the sake of cross-national comparability, one may have to accept the household per capita approach at the individual level (b), which resembles the way standard income distribution analysis is done.[9] It must be emphasized, though, that this approach yields the lowest inequality levels due to the implicit re-distribution process within households. Another point that should be made is

[8] That is, all individuals within the household completed the interview.

[9] One should note that children below respondent’s age (17 and over) are effectively excluded from this type of analysis.

that if one were to use the household-based household per capita approach, the inequality and poverty values would fall precisely between the two individual-based approaches.

Next, we examine the effect the choice of the aggregation unit has on subgroup indices. Here we focus solely on the two individual approaches, that is, the “pure” individual approach compared to the one where we aggregate wealth across all adult household members and reassign the resulting per capita value to each. There is a clear pattern for the Gini coefficient shown in Table 9: the youngest profit the most from the implicit redistribution process within the household, and this effect diminishes with increasing age and falls below average starting with those aged 35 and over. The picture for the HSCV is not as clear as for the Gini, but here again we find the highest "redistribution" effect in favor of the youngest population. Since this somewhat mixed picture may be induced by outliers, we repeat the analysis and top-code at the 99th percentile of the wealth distribution. Although inequality decreases, there is not much change in the Gini. For the HSCV the picture becomes much clearer and consistent across age groups. The youngest still benefit the most and the internal redistribution effects are below average for all starting at age 45. Using the top-coded data and a top-sensitive inequality measure, however, this effect is again diminished for the older population.

Decomposition of inequality by gender indicates that women profit more from within-household redistribution than men, and this effect is much stronger using a top-sensitive inequality measure. Figure 5 reveals that moving from an individual to a household per capita perspective “increases” the net worth of women from 69,000 to 81,000 euros, while men “lost” 15,000 euros on average to the redistribution process. However, this change is driven solely by married couples. Whereas both unmarried men and unmarried women gain about 10,000 euros, it is married women who profit most from household internal redistribution, moving from 85,000 to 100,000 euros. In other words, a household-based approach hides an existing gender wealth gap of approximately 50,000 euros. According to our inequality results based on the HSCV (Table 9), wealth inequality among women overall is reduced by internal redistribution by as much as 50%, while there is a major increase to be found among men. Once we repeat the analysis with top-coding we find that although women do benefit more than men from within-household distribution, the difference in the effect compared to men is only one percentage point for the Gini, but about eight percentage points for the HSCV.

5 Conclusions and future prospects

The goal of this paper is twofold. First, we want to provide an assessment of the effect of editing and multiple imputations in the 2002 SOEP. This was done in the aggregate and at individual level using descriptive and inequality measures. We find multiple imputation to be an effective means of coping with selective non-response. There is a significant impact of imputation on the share of wealth holders (leading to a 15% increase on average) and also on wealth components (30% increase on average). However, with respect to inequality we find ambiguous results. The Gini and HSCV for the wealth components are reduced for the whole population. For those with a component there is variation in the effect on the two inequality measures. We observe a reduction in inequality for owner-occupied property, financial assets, private pensions, tangibles, and other debts, and an increase for other property, business assets and property debt. Looking at the comprehensive outcome variable for the whole population—net worth—the Gini decreases, whereas the top-sensitive HSCV doubles. The non-random selectivity built into the missing process and the consideration of this selectivity in the imputation process clearly contribute to this finding.

A second goal is to provide an initial assessment of how the choice of unit of aggregation/unit of analysis affects the wealth distribution and inequality analysis. In accordance with our hypotheses, wealth measured in "household per capita" terms is less unequally distributed as a result of significant redistribution within households, also providing proof of the existing gender wealth gap.

With respect to future work on the production of cross-nationally comparative wealth information (such as LWS), one may want to develop a more comprehensive wealth measure by also considering (surveying or simulating) public pension entitlements. These may indeed turn out to be more relevant in some countries than in others, thus also exerting an effect on the level and structure of individuals' wealth portfolios. In any case, as has been shown above, the post-data-collection treatment of measurement errors, especially with respect to imputation of missing values due to non-response, will have an impact on cross-national comparability and may thus require some harmonization above and beyond the data collection itself. Last but not least, the significant differences among inequality measures resulting from the use of individual versus household-level data may require sacrificing the "superior" individual-level data for the sake of cross-national comparability.

In future research, we plan to compare SOEP-based results to results derived from the US Survey of Consumer Finances (SCF) in terms of the scope and correlates of missing values, as well as the resulting impact of multiple imputations.

6 References

Biewen, M. (2001): Item non-response and inequality measurement: Evidence from the German earnings distribution. *Allgemeines Statistisches Archiv* 85 409-425.

Böheim, R. and Jenkins, S. (2000): Do current income and annual income measures provide different pictures of Britain's income distribution?, Working Paper 2000-16, Institute for Social and Economic Research, University of Essex.

Frick, Joachim R. and Grabka, Markus M. (2005): Item-non-response on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility. *Allgemeines Statistisches Archiv,* 89(1): 49-60

Frick, Joachim R. and Grabka, Markus M. (2006): Item-non-response and Imputation of Labor Income in Panel Surveys: A Cross-National Comparison, Paper presented at the 29th General Conference of The International Association for Research in Income and Wealth (IARIW), Joensuu, Finland, August 20 – 26, 2006. http://www.iariw.org/papers/2006/FrickGrabka.PDF

Frick, Joachim R., Grabka, Markus M., and Marcus, Jan (2007): Wealth data in the German Socio-Economic Panel Study (SOEP). DIW Data Documentation, Berlin: DIW Berlin.

Haisken-DeNew, John P. and Frick, Joachim R. (2005): Desktop Companion to the German Socio-Economic Panel Study (GSOEP), Version 8.0 – Update to Wave 21, DIW Berlin.

Heckman, James J. (1979): Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153-161.

Kennickell, Arthur (1998): Multiple Imputation in the Survey of Consumer Finances, Working paper, http://www.federalreserve.gov/pubs/OSS/oss2/papers/impute98.pdf

Sierminska, Eva; Smeeding, Timothy M. and Brandolini, Andrea (2006): The Luxembourg Wealth Study – A Cross-Country Database for Household Wealth Research, *Journal of Economic Inequality,* vol. 4 no. 3, December 2006

Sierminska, Eva; Smeeding, Timothy M. and Brandolini, Andrea (2006): Comparing wealth distribution across rich countries: First results from the Luxembourg Wealth, *Luxembourg Wealth Study Working Paper No. 1,* August 2006

Spiess, Martin and Goebel, Jan (2005): On the effect of item nonresponse on the estimation of a two-panel-waves wage equation, *Allgemeines Statistisches Archiv*, 89(1), 63-74

Wagner, Gert G. et al (2001): Lebenslagen in Deutschland. Der erste Armuts- und Reichtumsbericht der Bundesregierung. Forschungsprojekt Einkommensverteilung und Einkommensmobilität. Bundesministerium für Arbeit Sozialordnung, Bonn.

Wagner, Gert G., Frick, Joachim R. and Schupp, Jürgen (2006): Enhancing the Power of Household Panel Studies – The Case of the German Socio-Economic Panel Study (SOEP). DIW Data Documentation No. 13, August 2006, Berlin: DIW Berlin. http://www.diw.de/deutsch/produkte/publikationen/datadoc/docs/diw_datadoc_2006-013.pdf
see also: http://www.ciqss.umontreal.ca/Longit/session3_paper1.html

7 Tables

Table 1
Incidence of item non-response: source for editing and imputation (population share affected, unweighted)

a) Basis: Total Population [1]

	Owner-occupied property	Other Property	Financial Assets	Private Pensions	Business Assets	Tangible Assets	Total Assets	Main Property Debt	Other Property Debt	Other Debt	Total Debt	Net Worth
Observed	86.9	97.7	86.5	80.6	93.4	91.8	67.6	88.2	93.8	93.3	83.6	63.4
Edited	4.0	0.4	0.0	0.0	-	-	5.4	2.3	0.2	-	3.6	5.5
Imputed	9.1	1.9	13.5	19.4	6.6	8.2	27.0	9.5	6.0	6.7	12.8	31.1
Total	100	100	100	100	100	100	100	100	100	100	100	100

n=23,892

b) Basis Population holding wealth/debt component [1]

	Owner-occupied property	Other Property	Financial Assets	Private Pensions	Business Assets	Tangible Assets	Total Assets	Main Property Debt	Other Property Debt	Other Debt	Total Debt	Net Worth
Observed	73.5	81.5	77.6	67.5	66.8	71.5	57.4	70.3	79.3	84.3	69.2	54.0
Edited	9.8	2.9	0.1	0.0	-	-	7.1	9.7	2.9	-	9.0	7.4
Imputed	16.8	15.6	22.3	32.5	33.2	28.5	35.5	20.0	17.8	15.7	21.9	38.7
Total	100	100	100	100	100	100	100	100	100	100	100	100
N of obs. (cumulated across 5 MI)	47,985	14,645	53,870	61,115	5,995	11,315	90,750	26,630	7,420	13,315	39,430	86,963
N of obs. (basic N)	9,597	2,929	10,774	12,307	1,247	2,263	18,185	5,326	1,484	2,663	7,886	17,393

Note: Calculations are based on flags that refer to total market value and not personal share.
[1] Individuals aged 17 and over.
Source: SOEP 2002; authors calculations.

Table 2
Population share holding wealth components: impact of imputation

	Observed [1]	Final [2]	% deviation [3]
Owner-occupied property	31.3	36.2	15.7
Other Property	8.2	10.0	22.0
Financial Assets	36.2	43.0	18.8
Private Pensions	35.1	36.2	3.1
Business Assets	2.7	4.2	55.6
Tangible Assets	5.8	8.4	44.8
Total Assets	63.2	73.5	16.3
Main Property Debt	15.1	18.2	20.5
Other Property Debt	3.7	4.6	24.3
Other Debt	9.1	10.7	17.6
Total Debt	40.0	46.2	15.5

Population: Adult population (17 years and over) with interview. Weighted.
[1] Only those with observed values are included.
[2] After editing and imputation.
[3] Calculated as (final-observed)/final.
Source: SOEP 2002; authors' calculations.

Table 3
The effect of imputation and editing on aggregate wealth in the 2002 SOEP
(individual level, weighted, in millions euro)

	Observed [1]	Final [2]	% Share imputed [3]
Owner-occupied property	2,408,607	3,421,346	29.6
Other Property	816,755	1,099,007	25.7
Financial Assets	500,767	670,272	25.3
Private Pensions	425,161	613,777	30.7
Business Assets	344,622	682,977	49.5
Tangible Assets	64,067	95,168	32.7
Total Assets	4,559,979	6,582,548	30.7
Main Property Debt	407,168	593,092	31.3
Other Property Debt	227,194	302,551	24.9
Other Debt	148,468	180,414	17.7
Total Debt	782,831	1,076,058	27.3
Net Worth	3,777,148	5,506,490	31.4

Note:
[1] For each component only those with observed personal share and value are included. Totals are based on the sum of components and not on individual data.
[2] After editing and imputation
[3] Share imputed is calculated as (final-observed)/final
Source: SOEP 2002; authors' calculations.

Table 4
The effect of editing and imputation on summary statistics of wealth components (individual, unweighted)

	Mean		% change[3]	Median		% change[3]
	Observed [1]	Final [2]		Observed [1]	Final [2]	
Owner-occupied property (PR)	38,008	57,660	51.7	0	0	0.0
se	(296)	(311)				
if PR>0	152,360	143,546	-5.8	125,000	117,313	-6.1
se	(734)	(579)				
Other property (IR)	14,847	23,493	58.2	0	0	0.0
se	(464)	(605)				
if IR>0	189,332	191,637	1.2	100,000	90,000	-10.0
se	(4,918)	(4,720)				
Financial Assets (TFA)	8,264	11,620	40.6	0	0	0.0
se	(159)	(156)				
if TFA>0	28,066	25,768	-8.2	10,000	10,000	0.0
se	(456)	(335)				
Private Pensions (PA)	8,105	11,772	45.2	0	0	0.0
se	(173)	(167)				
if PA>0	25,028	23,011	-8.1	10,000	9,980	-0.2
se	(455)	(320)				
Business Assets (BA)	8,325	18,223	118.9	0	0	0.0
se	(1,001)	(1,609)				
if BA>0	301,674	363,117	20.4	50,000	50,569	1.1
se	(31,939)	(31,907)				
Tangible Assets (DRCL)	1,273	2,002	57.3	0	0	0.0
se	(76)	(82)				
if DRCL>0	22,088	21,136	-4.3	8,000	9,492	18.6
se	(1,148)	(853)				
Main Property Debt (MG)	6,995	11,296	61.5	0	0	0.0
se	(90)	(106)				
if MG>0	53,108	50,673	-4.6	42,500	39,711	-6.6
se	(467)	(390)				
Other Property Debt (OMG)	4,598	7,333	59.5	0	0	0.0
se	(180)	(242)				
if OMG>0	113,780	118,064	3.8	56,000	58,110	3.8
se	(3,680)	(3,668)				
Other Debt (NHD)	2,096	2,981	42.2	0	0	0.0
se	(111)	(119)				
if NHD>0	26,545	26,744	0.8	9,650	10,000	3.6
se	(1,224)	(1,050)				

Note:
Standard errors in parentheses. Results are corrected for multiple imputation. Star (*) indicates means are significantly different at the 95% level.
[1] Only those with observed personal share and value are included. [2] After editing and imputation [3] (final-obs)/obs
Source: SOEP 2002; authors´ calculations.

Table 5
The effect of editing and imputation on summary statistics of wealth components: quantiles (weighted, individual level)

	Observed [1]	Edited	Imputed	Final [2]	% change[3]
Owner-occupied property (PR)					
1	0	0	0	0	-
5	0	32,500	0	0	-
10	0	50,000	0	0	-
25	0	71,250	0	0	-
50 (median)	0	102,500	76,565	0	-
75	60,000	137,500	128,547	90,000	50
90	150,000	188,750	200,000	153,388	2
95	200,000	219,000	253,040	215,000	8
99	400,000	325,000	419,815	400,000	0
n(MI) =	101,500	6,065	11,895	119,460	
Other property (IR)					
1	0	750	0	0	-
5	0	6,250	0	0	-
10	0	27,500	0	0	-
25	0	36,250	0	0	-
50 (median)	0	70,000	0	0	-
75	0	137,500	24,002	0	-
90	0	182,000	114,763	0	-
95	63,912	224,665	217,757	87,500	37
99	300,000	925,000	666,608	300,000	0
n(MI) =	110,785	490	8,185	119,460	
Financial Assets (TFA)					
1	0	0	0	0	-
5	0	0	0	0	-
10	0	0	0	0	-
25	0	0	2,500	0	-
50 (median)	0	49000	7,520	0	-
75	5,000	148500	16,415	9,000	80
90	20,000	181500	31,259	25,000	25
95	40,000	181500	48,140	45,000	13
99	125,000	181500	120,000	125,000	0
n(MI) =	99,900	50	19,510	119,460	
Private Pensions (PA)					
1	0	10000	0	0	-
5	0	10000	0	0	-
10	0	10000	0	0	-
25	0	10000	2,500	0	-
50 (median)	0	10000	5,982	0	-
75	5,000	10000	15,124	7,669	53
90	20,000	10000	32,905	23,736	19
95	40,000	10000	53,412	40,000	0
99	100,000	10000	128,203	102,000	2
n(MI) =	96,285	5	23,170	119,460	

Table 5 (continued)

	Observed [1]	Edited	Imputed	Final [2]	% [3]
Business Assets (BA)					
1	0	n.a	0	0	-
5	0	n.a	0	0	-
10	0	n.a	0	0	-
25	0	n.a	0	0	-
50 (median)	0	n.a	0	0	-
75	0	n.a	0	0	-
90	0	n.a	48,669	0	-
95	0	n.a	142,382	0	-
99	90,000	n.a	1,002,906	144,000	60
n(MI) =	111,530	-	7,930	119,460	
Tangible Assets (DRCL)					
1	0	n.a	0	0	-
5	0	n.a	0	0	-
10	0	n.a	0	0	-
25	0	n.a	0	0	-
50 (median)	0	n.a	0	0	-
75	0	n.a	4,787	0	-
90	0	n.a	13,877	0	-
95	3,500	n.a	22,142	6,000	71
99	20,000	n.a	60,486	28,463	42
n(MI) =	109,670	-	9,790	119,460	
Main Property Debt (MG)					
1	0	0	0	0	-
5	0	0	0	0	-
10	0	0	0	0	-
25	0	0	0	0	-
50 (median)	0	23,540	0	0	-
75	0	46,250	24,682	0	-
90	20,000	75,000	56,594	33,250	66
95	50,000	90,000	76,683	60,000	20
99	117,500	150,000	138,762	120,000	2
n(MI) =	103,115	4,270	12,075	119,460	
Other Property Debt (OMG)					
1	0	0	0	0	-
5	0	0	0	0	-
10	0	0	0	0	-
25	0	20,000	0	0	-
50 (median)	0	53,750	0	0	-
75	0	86,250	0	0	-
90	0	211,250	27,853	0	-
95	0	211,250	74,943	0	-
99	104,908	4,000,000	207,949	104,908	0
n(MI) =	111,835	290	7,335	119,460	
Other Debt (NHD)					
1	0	n.a	0	0	-
5	0	n.a	0	0	-
10	0	n.a	0	0	-
25	0	n.a	0	0	-
50 (median)	0	n.a	0	0	-
75	0	n.a	0	0	-
90	0	n.a	13,717	3,000	-
95	8,000	n.a	25,531	10,000	25
99	40,000	n.a	139,440	50,000	25
n(MI) =	111,410	-	8,050	119,460	

Note: [1] Only those with observed personal share and value are included. [2] After editing and imputation. [3] (final-obs)/obs
Source: SOEP 2002; authors' calculations.

Table 6
The effect of editing and imputation on wealth inequality (individual level, weighted)

	Total Population				Population with component		
	Observed [1]	Final [2]	% change [3]		Observed [1]	Final [2]	% change [3]
Owner-occupied property							
Gini	0.814	0.761	-6.5	*	0.345	0.341	-1.3
se	(0.002)	(0.003)			(0.004)	(0.003)	
HSCV	2.314	1.688	-27.0	*	0.298	0.293	-1.6
se	(0.064)	(0.059)			(0.015)	(0.019)	
Other property							
Gini	0.968	0.961	-0.7	*	0.590	0.608	3.1
se	(0.002)	(0.002)			(0.017)	(0.017)	
HSCV	43.643	38.848	-11.0		2.995	3.434	14.6
se	(9.555)	(5.395)			(0.744)	(0.511)	
Financial Assets							
Gini	0.871	0.833	-4.3	*	0.637	0.612	-4.0
se	(0.005)	(0.005)			(0.009)	(0.008)	
HSCV	12.167	8.861	-27.2	*	4.019	3.527	-12.2
se	(2.380)	(1.659)			(0.817)	(0.689)	
Private Pensions							
Gini	0.869	0.832	-4.2	*	0.656	0.644	-1.9
se	(0.005)	(0.004)			(0.010)	(0.007)	
HSCV	12.083	8.309	-31.2	*	4.302	3.658	-15.0
se	(3.339)	(2.095)			(1.278)	(0.984)	
Business Assets							
Gini	0.994	0.993	-0.1	*	0.783	0.825	5.5
se	(0.001)	(0.001)			(0.037)	(0.019)	
HSCV	938.845	627.704	-33.1		24.884	25.245	1.4
se	(414.253)	(130.341)			(11.360)	(5.199)	
Tangible Assets							
Gini	0.977	0.966	-1.1	*	0.626	0.598	-4.5
se	(0.001)	(0.001)			(0.015)	(0.012)	
HSCV	133.013	77.029	-42.1	*	7.812	6.043	-22.6
se	(60.300)	(32.056)			(3.944)	(2.779)	
Main Property Debt							
Gini	0.925	0.899	-2.8	*	0.440	0.445	1.0
se	(0.003)	(0.003)			(0.011)	(0.010)	
HSCV	6.671	5.067	-24.1	*	0.464	0.514	10.8
se	(0.603)	(0.629)			(0.075)	(0.103)	
Other Property Debt							
Gini	0.984	0.981	-0.3	*	0.579	0.586	1.1
se	(0.001)	(0.001)			(0.020)	(0.019)	
HSCV	79.107	74.775	-5.5		2.493	2.935	17.7
se	(17.897)	(11.475)			(0.649)	(0.488)	
Other Debt							
Gini	0.969	0.965	-0.4	*	0.683	0.674	-1.4
se	(0.003)	(0.003)			(0.037)	(0.032)	
HSCV	100.788	77.855	-22.8	*	9.250	7.845	-15.2
se	(42.403)	(30.897)			(4.128)	(3.332)	

Note:
Standard errors in parentheses. Star (*) indicates Gini values are significantly different at the 95% level.
[1] Only those with observed personal share and value are included. [2] After editing and imputation
[3] (final-obs)/obs
Source: SOEP 2002; authors' calculations.

Table 7
Asset poverty and the effect of editing and imputation on overall wealth (net worth, individual level)

	Observed[1]	**Final[2]**	**% change[3]**
Mean	60,235	81,713	35.7
se	(819)	(1,302)	(58.9)
Mean if NW>0	104,466	117,812	12.8
se	(1,039)	(1,539)	
1	-20,000	-20,000	0.0
5	-3,118	-1,540	-50.6
10	0	0	
25	0	0	
50 (median)	5,000	15,000	200.0
75	60,000	96,588	61.0
90	174,760	208,000	19.0
95	275,000	313,942	14.2
99	600,000	729,711	21.6
Gini	0.837	0.787	-6.0
HSCV	6.791	14.681	116.2
Headcount ratio: FGT(0)	0.451	0.427	-5.4
Average normalized poverty gap: FGT(1)	0.828	0.506	-38.9
Average squared normalized poverty gap: FGT(2)	14.929	9.216	-38.3
N of obs. (cumulated across 5 MI)	73,450	115,675	
N of obs. (basic N)	14,690	23,135	

Note: Standard errors in parentheses. The asset poverty threshold is set at half the median of net worth.
[1] Only those with observed personal share and value are included. [2] After editing and imputation
[3] (final-obs)/obs
Source: SOEP 2002; authors' calculations.

Table 8
The effect of the choice of the aggregation unit on distribution statistics (net worth and post-government income)

Unit of Analysis	Net worth					Disposable household income per adult		
	Individual [1]		Household [2]	% deviation		Individual [1]	Household [2]	
	Individual (a)	Household per capita (b)	Household per capita (c)	(b)-(a) / (a)	(c)-(b) / (b)	Household per capita (B)	Household per capita (C)	% deviation (C)-(B) / (B)
Mean	81,713	81,797	80,011	0.1	-2.2	16,175	16,684	3.1
se	(1,302)	(1,003)	(1,547)					
1	-20,000	-20,000	-20,251	0.0	1.3			
5	-1,540	-2,360	-2,646	53.2	12.1			
10	0	0	0	0.0	0.0	7,530	7,439	-1.2
25	0	1,897	317	0.0	-83.3	10,182	10,264	0.8
50 (median)	15,000	27,500	20,000	83.3	-27.3	13,932	14,359	3.1
75	96,588	99,000	93,418	2.5	-5.6	19,430	20,202	4.0
90	208,000	197,081	198,000	-5.2	0.5	26,289	27,199	3.5
95	313,942	290,981	298,347	-7.3	2.5	32,345	33,705	4.2
99	729,711	612,148	632,441	-16.1	3.3	51,294	53,168	3.7
min	-3,692,144	-1,152,392	-1,152,392	-68.8	0.0			
max	99,221,992	51,763,632	51,763,632	-47.8	0.0			
sum	5.506E+12	5.512E+12	5.642E+12	0.1	2.4			
Inter quartile range (75:25)	96,588	97,103	93,101	0.5	-4.1	9,248	9,938	7.5
Gini	0.787	0.734	0.762	-6.8	3.9	0.292	0.301	3.4
se	(0.006)	(0.006)	(0.007)					
HSCV	14.681	8.695	11.496	-40.8	32.2	0.257	0.303	17.8
se	(4.575)	(2.493)	(3.848)					
Headcount ratio: FGT(0)	0.427	0.413	0.409	-3.1	-1.0	0.140	0.156	11.8
Average normalized poverty gap: FGT(1)	0.506	0.396	0.430	-21.8	8.7	0.035	0.041	19.4
Average squared normalized poverty gap: FGT(2)	9.216	1.065	1.491	-88.4	40.0	0.015	0.018	26.6
N of obs. (cumulated across 5 MI)	115,675	115,675	61,540					
N of obs. (basic N)	23,135	23,135	12,308					

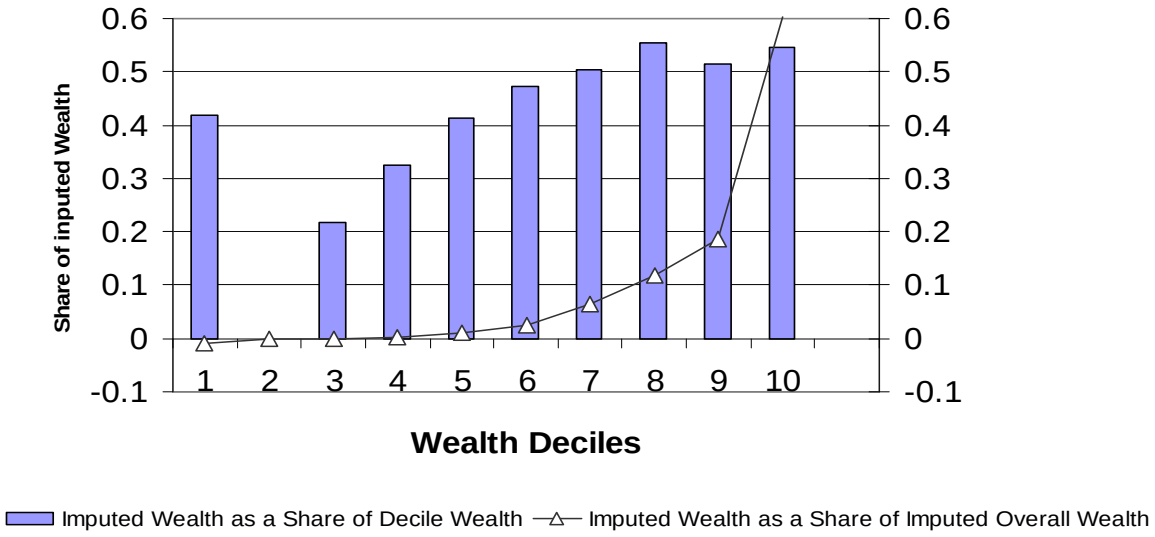
Note: Standard errors in parentheses. The asset poverty threshold is set at half the median of net worth.
[1] Data is based on all individuals that have completed the interview
[2] Data is based on all households that have completed the interview. The household may include individuals that do not have completed interviews.
Source: SOEP 2002; authors' calculations.

Table 9
The effect of the choice of the aggregation unit and top coding on subgroup indices (net worth, weighted)

	Individual [1]			Individual [1] (Top coded at 1%)		
Unit of Analysis	Individual	Household per capita	%deviation (HH - Ind)/Ind	Individual	Household per capita	%deviation (HH - Ind)/Ind
TOTAL						
Gini	0.787	0.734	-6.8	0.756	0.697	-7.8
se	(0.006)	(0.006)		(0.005)	(0.004)	
HSCV	14.681	8.695	-40.8	1.598	1.195	-25.2
se	(4.575)	(2.493)		(0.043)	(0.018)	
By AGE:						
Gini						
<=24	0.974	0.758	-22.2	0.968	0.746	-23.0
25-34	0.975	0.860	-11.7	0.946	0.849	-10.3
35-44	0.792	0.756	-4.5	0.754	0.708	-6.0
45-54	0.740	0.709	-4.2	0.684	0.654	-4.5
55-64	0.693	0.669	-3.4	0.645	0.617	-4.4
65-74	0.680	0.654	-3.8	0.656	0.628	-4.2
75+	0.721	0.688	-4.6	0.700	0.666	-4.9
HSCV						
<=24	80.403	3.129	-96.1	17.560	1.653	-90.6
25-34	5.872	8.273	40.9	3.876	2.353	-39.3
35-44	26.602	22.291	-16.2	1.790	1.269	-29.1
45-54	23.330	12.136	-48.0	1.113	1.017	-8.6
55-64	4.027	4.064	0.9	0.925	0.813	-12.1
65-74	2.087	1.519	-27.2	0.964	0.836	-13.3
75+	2.291	1.816	-20.7	1.172	0.967	-17.4
By GENDER:						
Gini						
Female	0.788	0.728	-7.6	0.766	0.704	-8.1
Male	0.782	0.738	-5.6	0.743	0.689	-7.3
HSCV						
Female	20.117	9.232	-54.1	1.727	1.235	-28.5
Male	3.404	8.163	139.8	1.456	1.151	-20.9

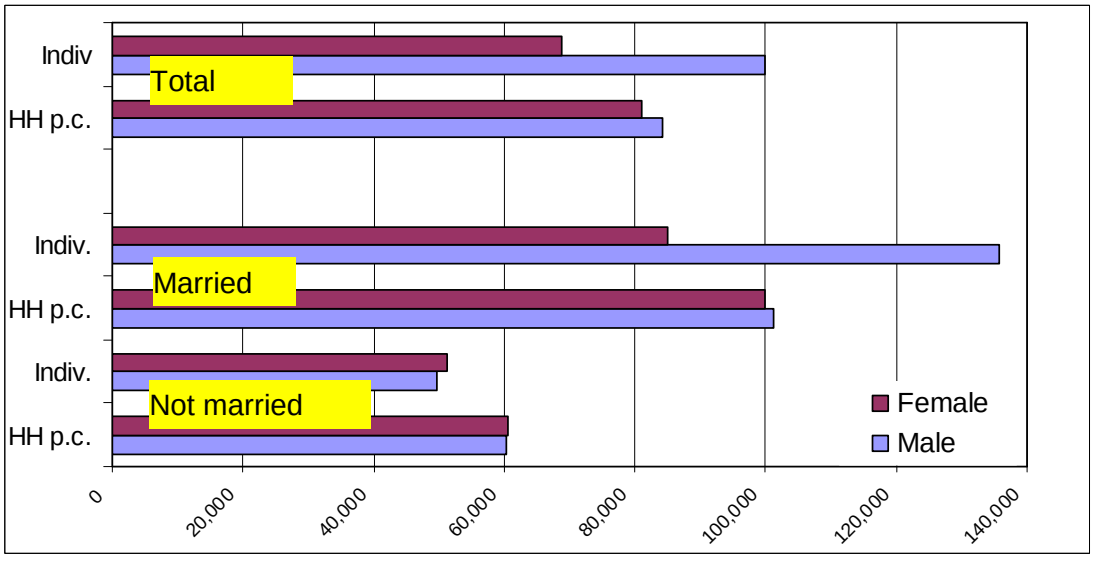
Note: Bootstrap standard errors (100 reps).
[1] Data is based on all individuals that have completed the interview (n=23135)
Source: SOEP 2002; authors' calculations

Figure 4: Share of wealth and decile wealth imputed by deciles



Source: SOEP 2002; authors' calculations.

Figure 5: Wealth from the individual vs. the household perspective: "gender wealth gap"



Source: SOEP 2002; authors' calculations.

8 Appendices

Appendix 1
Wealth questions in the 2002 SOEP individual questionnaire:

The following questions are under the subtitle of "Your personal assets and liabilities"

The accumulation of wealth is an important subject in all levels of society nowadays, especially in view of the future for provisions for old age and the reform of pensions. The German Institute for Economic Research (DIW) in Berlin is currently undertaking a large research project in this field. It is aiming to produce an accurate picture of the financial circumstances of the nation's citizens. We would like to invite you to take part in this project. In order to do so, together with you, we would like to create your own personal "assets and liabilities statement", which may also be able to help you gain a better view of your finances. You can be absolutely sure that your details will be handled with confidentiality and will only be used for economic evaluation.

+ *Please continue to questions* **A** *to* **G** *on the following pages.*

If you have no information regarding to these questions, then please proceed to question 86

Your personal assets and liabilities statement

Do you possess the following types of property or wealth?
If yes, then please estimate its current value.

(A) Are you personally the owner of the house or apartment in which you live?

Yes .... ☐ ⇨
No ..... ☐ ⇩

Value:
If you were to sell today, how much would you receive for your house/apartment including land? EURO ☐

Burden:
If you still have a loan taken out on your house/apartment, how high is the remaining debt (excluding interest)? EURO ☐

Personal share of property:
Are you the sole owner (100%) or co-owner (e.g. with your spouse)? Sole Owner ☐
If the latter, how high is your own share? Share in % ☐☐

(B) Apart from the property you live in, do you possess another house or more land?

Yes .... ☐ ⇨
No ..... ☐ ⇩

Type and number of properties:

What type of property is it?

- One family house / Free hold flat (not used by yourself) ..... ☐
- Multiple family house / Apartment house ..... ☐
- Holiday home / Weekend home ..... ☐
- Undeveloped land ..... ☐
- Other property ..... ☐

How many of this type of property do you have ? (excluding the one used by yourself) Amount ..... ☐☐

Value:
If you were to sell your property today (excluding the one used by yourself), how much would you receive? EURO ☐

Personal share of property:
Are you the sole owner (100%) or co-owner (e.g. with your spouse)? Sole Owner ☐
If the latter, how high is your own share? Share in % ☐☐

Burden:
If you still have a loan taken out on your property, how high is the remaining debt (excluding interest)? EURO ☐

(C) **Do you possess financial assets over the value of 2500 EURO in the form of a savings balance, savings bonds, bonds, shares or investments?**

Yes .... ☐ ⇨
No ..... ☐ ⇩

Value:
How high do you estimate the value of your financial assets? EURO ☐

Personal share of property:
Are these financial assets in your name or do they stretch over joint accounts, i.e. with your spouse? Sole Owner ☐

If the latter, how high is your share? Share in % ☐☐

(D) **Do you currently possess life insurance or a private pension plan or a buildings savings account?**

Yes .... ☐ ⇨
No ..... ☐ ⇩

Value:
How high do you estimate the cash surrender value of these policies or financial assets to be? EURO ☐

(E) **Are you the owner of a commercial enterprise, i.e. a company, a shop, an office, a practice or an agricultural enterprise, or are you involved in an enterprise such as the above foremetioned?**

Yes .... ☐ ⇨
No ..... ☐ ⇩

Personal share of property:
Are you the sole owner or co-owner of this enterprise, e.g. GBR, GmbH or KG? Sole Owner ☐ Co-owner ☐

Value:
How high do you estimate the current value of your enterprise or of your share to be?
This is the price before tax, which you would receive at the sale of your enterprise or your share, taking into account any remaining financial burdens. EURO ☐

(F) **Do you possess any tangible assets over 2,500 EURO (excluding motor vehicles) in the form of gold, jewellery, coins or valuable collections?**

Yes .... ☐ ⇨
No ..... ☐ ⇩

Value:
If it were possible to estimate the value of these tangible assets: How high would the total value be? EURO ☐

(G) Leaving aside any mortgages on house or property or house-building loan:
Do you at the present time have any debts relating to credit that you as a private individual have taken on at a bank or a similar institution or a another individual, for which you are accountable?
This is limited to debts greater than 2,500 EURO. This does not include mortgages or house-building loans!

Yes .... ☐ ⇨
No ..... ☐ ⇩

Burden:
Current remaining debt (without interest): EURO ☐

Question ***86***
next page!

Appendix 2
Comparison of total wealth of private households with national balance sheet 2002 (in billion euros)

	National balance sheet (1)	SOEP[1] (2)	(2) / (1) in %
Gross wealth (excluding durables)	9,025	6,493	71.9
Property	4,640	4,526	97.5
Financial assets I	3,730	1,284	34.4
Financial assets II [2]	(2.630)	(1.284)	(48,8)
Net business assets [3]	655	683	104.3
Liabilities [4]	1,206	1,119	92.8
Mortgages [4]	1,002	939	93.7
Other debts [4 5]	204	180	88.2
Net Wealth (excluding durables)	7,819	5,374	68.7
Net Wealth (excluding durables, based on financial assets II)	6,719	5,374	80.0
Durables [6]	968	95	9.8

1: Sub-samples A-G, imputed wealth information.

2: Without non-profit-institutions and without currency and transferable deposits, certain claims on insurance corporations (for example health insurance and private pension funds) as well as claims from company pension commitments all of which are not covered by SOEP-microdata.

3: Ammermüller et al. (2005), Table 54, p.84.

4: Nikolaus Bartzsch and Elmar Stöss (2006): Measuring German household debt: Financial accounts data and disaggregated survey data as complementary statistics. Financial accounts of the Deutsche Bundesbank (without entrepreneurial loans). Table 10: Financial assets and debt of German households. Prepared for the IFC conference in Basle, August 2006.

5: For commercial and consumption purposes.

6: National balance sheets include all personal belongings (Ammermüller et al. 2005, p. 100).

Source: Andreas Ammermüller, Andrea M. Weber and Peter Westerheide (2005): Die Entwicklung und Verteilung des Vermögens privater Haushalte unter besonderer Berücksichtigung des Produktivvermögen. Abschlussbericht zum Forschungsauftrag des Bundesministeriums für Gesundheit und Soziale Sicherung. Zentrum für europäische Wirtschaftsforschung (ZEW), p. 101.

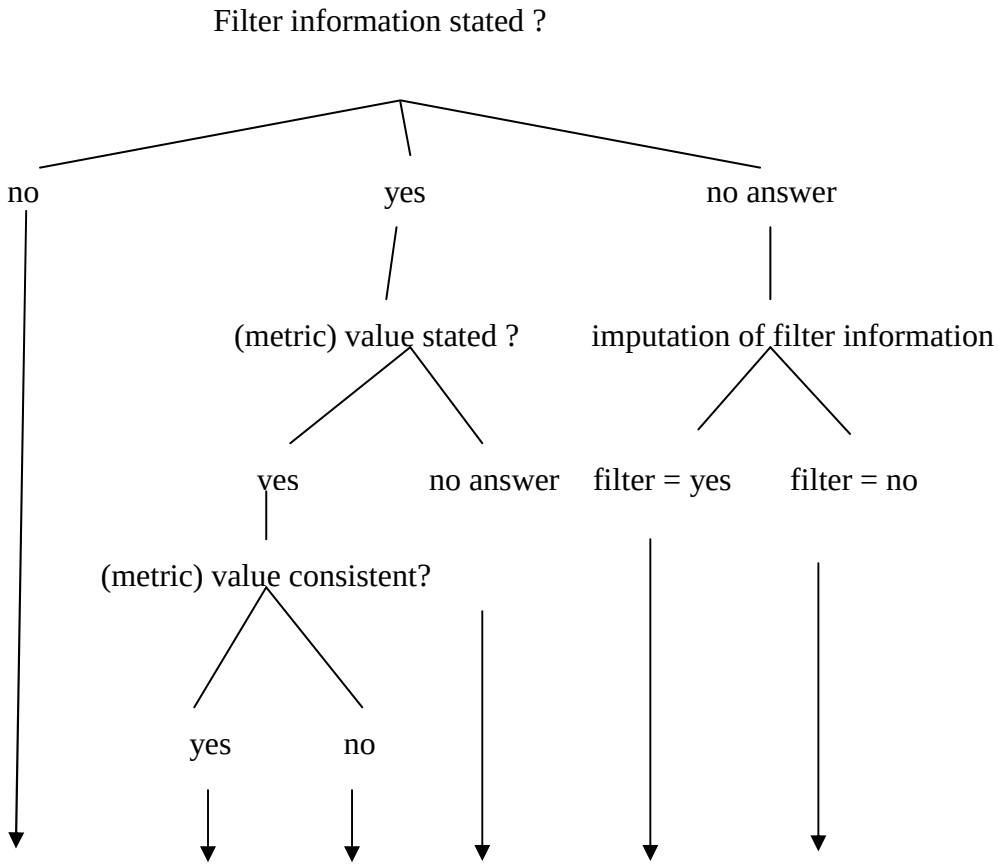
Appendix 3
Estimating the probability for total gross wealth (TA) and total debt (TD) to be affected by „imputation“ (due to item non-response) or by “editing” (because of inconsistency): Probit model with Heckman sample selection correction

	Imputed TA (1) Model A		Imputed or edited TA (2) Model B		Imputed TD (3) Model A		Imputed or edited TD (4) Model B	
Variable	**coeff**	**std.dev.**	**coeff**	**std.dev.**	**coeff**	**std.dev.**	**coeff**	**std.dev.**
Age	-.0012	(.0050)	.0040	(.0048)	-.0086	(.0093)	-.0194**	(.0083)
Age (squared)	.0014	(.0470)	-.0450	(.0456)	.0748	(.0840)	.1486**	(.0754)
Male	-.0640***	(.0198)	-.0648***	(.0192)	-.1266***	(.0266)	-.0969***	(.0241)
Migrant	.0646	(.0396)	.1088***	(.0377)	.0463	(.0515)	.0449	(.0470)
Low education	-.0288	(.0342)	.0852***	(.0324)	.0077	(.0439)	.0828**	(.0395)
High education	-.0894***	(.0252)	-.0990***	(.0243)	-.1398***	(.0329)	-.1353***	(.0297)
Children in household	-.0566**	(.0243)	.0278	(.0235)	.0372	(.0368)	.0496	(.0336)
Lived in	.0455*	(.0237)	-.0002	(.0231)	.0881**	(.0352)	.1229***	(.0316)
Self employed	.2251***	(.0376)	.1336***	(.0371)	.0674	(.0474)	-.0339	(.0435)
Civil servant	-.0572	(.0441)	-.0761*	(.0424)	-.0658	(.0560)	-.0932*	(.0503)
Number of interviews	-.0092***	(.0021)	-.0115***	(.0020)	-.0036	(.0028)	-.0025	(.0026)
Interview mode:	.0338	(.0266)	.0882***	(.0257)	-.0333	(.0356)	.0586*	(.0325)
Interview mode: CAPI	.6174***	(.0261)	.6492***	(.0266)	.1163***	(.0352)	.1561***	(.0325)
Interview mode: Postal	.3556***	(.0362)	.4085***	(.0357)	.3233***	(.0482)	.3596***	(.0453)
Constant	-.4557***	(.1438)	-.3304**	(.1392	-.4431	(.2910	.2259	(.2564
	TA=yes		TA=yes		TD=yes		TD=yes	
Age	.0567***	(.0038)	.0553***	(.0038)	.1054***	(.0038)	.1049***	(.0038)
Age (squared)	-.4292***	(.0397)	-.4155***	(.0395)	-.8413***	(.0392)	-.8382***	(.0390)
Male	.0794***	(.0209)	.0782***	(.0208)	.0814***	(.0186)	.0808***	(.0186)
Married	.4558***	(.0253)	.4654***	(.0250)	.5262***	(.0220)	.5359***	(.0217)
Migrant	-.6625***	(.0295)	-.6668***	(.0295)	-.4576***	(.0287)	-.4582***	(.0287)
Low education	-.4466***	(.0252)	-.4465***	(.0251)	-.2909***	(.0252)	-.2920***	(.0252)
High education	.1587***	(.0323)	.1641***	(.0322)	.0385	(.0246)	.0410*	(.0246)
Children in household (< 14 yrs)	.0406	(.0262)	.0370	(.0261)	.2777***	(.0238)	.2728***	(.0238)
Rural area	.2554***	(.0275)	.2562***	(.0273)	.3543***	(.0240)	.3482***	(.0238)
Metropolitan area	-.0956***	(.0310)	-.1023***	(.0307)	-.2831***	(.0280)	-.2850***	(.0276)
Lived in East Germany in 1989	-.0635**	(.0271)	-.0725***	(.0271)	-.4106***	(.0245)	-.4094***	(.0244)
Unemployed	-.5596***	(.0418)	-.5657***	(.0413)	-.2411***	(.0411)	-.2412***	(.0405)
Self employed	.3124***	(.0597)	.3078***	(.0594)	.2709***	(.0405)	.2740***	(.0404)
Pensioner	-.1209***	(.0444)	-.1338***	(.0439)	-.1567***	(.0370)	-.1514***	(.0365)
Not employed (out of labor	-.3404***	(.0307)	-.3474***	(.0302)	-.2354***	(.0306)	-.2285***	(.0302)
Civil servant	.3493***	(.0740)	.3414***	(.0738)	.2723***	(.0471)	.2705***	(.0470)
Number of Interviews	.0019	(.0024)	.0022	(.0024)	-.0038*	(.0020)	-.0037*	(.0020)
Interview mode: self admin.	.1560***	(.0272)	.1554***	(.0271)	.1514***	(.0240)	.1488***	(.0240)
Interview mode: CAPI	.1601***	(.0264)	.1617***	(.0263)	.0847***	(.0237)	.0817***	(.0237)
Interview mode: Postal	.1890***	(.0370)	.1911***	(.0369)	.1734***	(.0332)	.1712***	(.0332)
Equivalent Pre-Gov't Income	.0126***	(.0006)	.0123***	(.0006)	.0056***	(.0004)	.0055***	(.0004)
Constant	1.2486***	(.0924)	-1.2074***	(.0922)	-3.3137***	(.0953)	-3.300***	(.0950)
/athrho	.0272	(.0709)	-.3468***	(.0684)	-.1970***	(.0691)	-.4782***	(.0641)
LR test on indep. equations	0.15		25.43***		7.77***		52.52***	
Number of obs	23892		23892		23892		23892	
Censored obs	5707		5707		11897		11897	
Uncensored obs	18185		18185		11995		11995	
Wald chi2(14)	895.83***		990.70***		137.02***		198.08***	
Log lik (full model)	-21200.08		-21708.02		-18865.85		-19813.89	

Note: *** sig. at 1% level; ** sig. at 5% level; * sig. at 10%level.
Model A looks only at the probability of Item non-response (and the consequential need for imputation), Model B analyses correlates of both types of measurement error (Item non-response as well as inconsistency).

Source: SOEP 2002, authors' calculations.

Appendix 4: Simplified decision tree for imputation and editing of wealth components



Component	Value=0 %	No change %	Editing %	Impute %	Impute %	Value=0 %	Total %
owner-occupied property	57,4	29,5	4,0	4,6	2,1	2,4	100
debt owner-occupied property	57,4	30,9	2,3	5,0	2,1	2,4	100
other property	83,4	9,5	0,4	1,9	0,5	4,4	100
debt other property	83,4	10,4	0,2	1,2	0,5	4,4	100
financial assets	51,5	35,0	0,0	7,6	2,5	3,4	100
private pension	45,8	34,8	0,0	13,8	2,9	2,7	100
business assets	89,9	3,5	0,0	1,5	0,3	4,9	100
tangible assets	85,0	6,8	0,0	2,5	0,2	5,5	100
consumer debts	83,9	9,4	0,0	1,5	0,2	5,0	100