

Frick, Joachim R.; Grabka, Markus M.; Marcus, Jan

Working Paper

Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP)

SOEPPapers on Multidisciplinary Panel Data Research, No. 18

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Frick, Joachim R.; Grabka, Markus M.; Marcus, Jan (2007) : Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP), SOEPPapers on Multidisciplinary Panel Data Research, No. 18, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/150567>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

SOEPpapers
on Multidisciplinary
Panel Data Research

18

Joachim R. Frick
Markus M. Grabka
Jan Marcus



SOEP

DIW Berlin

German Institute
for Economic Research

The German
Socio-Economic
Panel Study

**Editing and Multiple Imputation of Item-Non-Response
in the 2002 Wealth Module of the
German Socio-Economic Panel (SOEP)**

Berlin, April 2007

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

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

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

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

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

Editors:

Georg **Meran** (Vice President DIW Berlin)
Gert G. **Wagner** (Social Sciences)
Joachim R. **Frick** (Empirical Economics)
Jürgen **Schupp** (Sociology)
Conchita **D'Ambrosio** (Welfare Economics)
Christoph **Breuer** (Sport Science, DIW Research Professor)
Anita I. **Drever** (Geography)
Frieder R. **Lang** (Psychology, DIW Research Professor)
Jörg-Peter **Schräpler** (Survey Methodology)
C. Katharina **Spieß** (Educational Science)
Martin **Spieß** (Statistical Modelling)
Viktor **Steiner** (Public Economics, Department Head DIW Berlin)
Alan S. **Zuckerman** (Political Science, DIW Research Professor)

ISSN: 1864-6689

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Koenigin-Luise-Strasse 5
14195 Berlin, Germany
Contact: Uta Rahmann | urahmann@diw.de

Joachim R. Frick*, Markus M. Grabka* and Jan Marcus**

**Editing and Multiple Imputation of Item-Non-Response
in the 2002 Wealth Module of the
German Socio-Economic Panel (SOEP)**

Berlin, March 2007

* DIW Berlin, Department German Socio-Economic Panel,(SOEP), jfrick@diw.de,
mgrabka@diw.de

** University of Konstanz, jan.marcus@uni-konstanz.de

Contents

1 Introduction	1
2 Principles of editing and imputation.....	3
3 Owner-occupied property.....	6
3.1 Consistency check.....	6
3.1.1 Filter of owner-occupied property	7
3.1.2 Market and debt value	7
3.1.3 Share of owner-occupied property	8
3.1.4 Consistency with mortgage and interest payments from the household questionnaire.....	9
3.1.5 Implausible values	9
3.2 Logical imputations	10
3.2.1 Imputation of missing filter information on owner-occupied property.....	10
3.2.2 Imputation of missing market values and outstanding mortgage debt.....	11
3.2.3 Imputation of missing personal shares of own property.....	11
3.3 Regression-based multiple imputation of market value and level of outstanding mortgage debt for owner-occupied property.....	12
4 Other property.....	17
4.1 Internal checks of consistency and logical imputations.....	18
4.2 Imputation.....	20
4.2.1 Imputation of the filter variable.....	20
4.2.2 Imputation of the personal share	20
4.2.3 Imputation of market value and debts	20
5 Financial assets	21
5.1 Checks of consistency.....	22
5.2 Logical imputations for co-owners	22
5.3 Imputation.....	23
5.3.1 Imputation of the filter variable.....	23
5.3.2 Imputation of the personal share	23
5.3.3 Imputation of the value of the financial assets	23

6 Private pension assets, business assets, tangible assets, debt from consumer credits.....	24
6.1 Checks of consistency/logical imputations.....	25
6.2 Imputation.....	25
6.2.1 Imputation of the filter variable.....	25
6.2.2 Imputation of the personal share.....	25
6.2.3 Imputation of the value.....	25
7 Restrictions of the 2002 SOEP wealth information.....	26
8 Comparison with the national aggregate information.....	26
9 Impact of imputation on wealth inequality.....	28
10 Conventions on naming variables.....	31
10.1 Variable list at the individual level.....	33
10.2 Variable list at the household level.....	36
10.3 How to perform analyses using multiply imputed values?.....	37
References.....	39
Appendix.....	40

List of figures and tables

Figure 1: Simplified decision tree for imputing and editing wealth components	5
Figure 2: Comparison of different imputation methods to the observed cases of market value of own property	13
Figure 3: Predicted market values with and without randomly chosen residuals	14
Figure 4: Comparison the distribution of imputates for cases with non-response (“unobserved predictions”) and for imputates for observed cases (“observed predictions”) with the distribution of truly observed cases	16
Figure 5: Variable naming conventions (file PWEALTH)	32
Table 1: Editing and imputation process for “other property” (co-ownership households only)	18
Table 2: Logical imputation/editing of financial assets	22
Table 3: Comparison of total wealth of private households with the 2002 national balance sheet	27
Table 4: Illustration of the different aggregation levels (using information on Financial Assets)	33
Table 5: Comparison of the applied imputation imputates (individual level)	40
Table 6: Comparison of the applied imputation imputates (individual level, for those owning the respective wealth component)	42
Table 7: Comparison of the applied imputation imputates (household level)	44
Table 8: Comparison of the applied imputation imputates (household level; only those owning the respective wealth component)	46
Table 9: Comparison of observed values and values edited/imputed by Frick/Grabka/Marcus (2007) and Schupp/Schaefer (2006) (unweighted results)	48
Table 10: Influence of the imputation/editing process (individual level, weighted)	49
Table 11: Influence of the imputation/editing process (household level, weighted)	50
Table 12: The effect of editing and imputation on wealth inequality (individual level, weighted)	51
Table 13: Set-up of and covariates used in regression model	52

1 Introduction

Since the mid-1980s there have been no consistent, complete microdata available on the wealth of private households in Germany, in particular on their private business equity. Furthermore, there is a lack of systematically collected data on the wealth of high-income earners.¹ The 2002 German Socio-Economic Panel (SOEP) attempted to fill this gap by collecting information on private household wealth, providing new insight into this important issue. One novelty of the 2002 SOEP questionnaire was the change in the unit of observation. In contrast to previous studies in Germany and elsewhere, SOEP did not collect this information at the household level but at the individual level (another exception to this general rule is the BHPS, see Taylor et al. 1998). In contrast to the 1988 SOEP wealth questionnaire, which was part of the household questionnaire, from 2002 on wealth questions were included in the standard individual questionnaire². Thanks to this change of observation unit, it is now possible to analyze the distribution of assets and liabilities not only at the household level but also at the individual level (see Frick, Grabka and Sierminska 2007), and thus also to look at the wealth distribution within households or between spouses or partners.

The 2002 SOEP questionnaire surveys seven components of wealth. These include information on owner-occupied housing (including mortgage debt), other property (including mortgage debt), financial assets, business assets, tangible assets, private pensions (including life insurance) and consumer credits. One shortcoming is the lack of information on pension entitlements through both company pensions and the statutory German social pension fund (“Gesetzliche Rentenversicherung” for blue-collar and white-collar workers as well as the pension entitlements for civil servants), due largely to the difficulty of obtaining data on pension entitlements for individuals still in the labor force.

¹ The EVS (“Einkommens- und Verbrauchsstichprobe”) of the Federal Statistical Office in Germany is the only survey that regularly collects wealth information of private households. However, the EVS has several disadvantages, such as non-coverage of business assets (after 1983) and inadequate coverage of the foreign population. Even more important, for wealth analyses, there is an upper income threshold effectively excluding high-income households from the sampling frame. For a comprehensive comparison of EVS and SOEP, see Becker et al. 2002.

² The wealth module consisted of two pages in the questionnaire, sequenced after many other income related questions. The general framework and topical module of the 2002 questionnaire focused on social security, which was thought to increase people’s awareness of the relevance of this topic. The complete 2002 individual questionnaire can be downloaded from the SOEP website http://www.diw.de/deutsch/sop/service/fragen/fr2002/fr_personen_en.pdf.

Like other population surveys, SOEP is affected by measurement error. This is especially true for questions on wealth. A typical type of measurement error is item non-response (INR), i.e., the failure to collect complete information on a specific item. Partial unit-non-response (PUNR) occurs in household surveys like SOEP when one or more members of a multi-person household do not take part in the survey while the rest do. An aggregation of wealth holdings across all members of a given household presumably leads to underestimation in the case of PUNR. Another problem arises from inconsistent information provided by members of the same household or couples sharing a specific wealth component: for example, couples who co-own their home. Here, the SOEP questionnaire asks for (an estimate of) the current market value of the home as well as the percentage share thereof owned by that individual. As such, the market value estimated by each of the two partners should coincide. Secondly, if the two partners are sole owners of the property, their respective shares should add up to 100%. Any deviation from this must be considered measurement error and corrected through some form of “editing” as opposed to “imputation”, which is used for missing information due to item non-response.

The aim of this paper is to describe the two strategies currently used by the SOEP group for handling these measurement errors:

1. Editing in the case of inconsistent data and the possibility for logical imputation.
2. Multiple imputation using Hock-Deck regression methods in the case of item non-response or partial unit non-response.

A preliminary study on the imputation of missing wealth information in the 2002 SOEP questionnaire was conducted by Schäfer and Schupp (2006). This study performed mean-based imputations for those observations which were known to hold a given wealth component.³ The approach described in the present paper can be seen as an improvement over the previous procedure since it solves all the non-response problems (INR and PUNR) associated with the 2002 SOEP wealth questionnaire by means of imputation. The state-of-the-art imputation techniques used here do not have the drawback of mean-based imputations, which typically

³ That is, the respective filter information was non-missing.

understate *true* variance⁴. A further improvement is the application of multiple imputation techniques here, in contrast to the single imputation performed by Schäfer and Schupp (2006).

This paper is structured as follows: Section 2 describes the principles used for the editing and imputation of SOEP wealth data. Sections 3 to 6 discuss how measurement errors are handled in the respective wealth components. Section 7 describes some general limitations of the SOEP wealth data. Section 8 compares the resulting wealth aggregates with corresponding information from official national statistics (SNA and *Bundesbank*). Section 9 presents the effects of editing and imputation on various measures of wealth inequality. Finally, Section 10 provides external researchers using SOEP data with information on the structure of the wealth data and hints for the proper use of multiply imputed data.

2 Principles of editing and imputation

The first step in handling measurement errors in the 2002 SOEP wealth questionnaire is to check for the consistency and plausibility of information across household members. Information is harmonized, i.e. edited, using specific rules which will be presented in the following sections. In our context, *editing* means changing a non-missing value into a new value (possibly including values of zero), while regression-based imputation is used for non-response. For selected components, however, imputation may be carried out by means of logical imputation, e.g., in the case of co-owner couples with one partner providing plausible information on his/her share of the wealth and the other providing none. In general, logical imputation involves a non-regression-based imputation derived from survey information given in the household questionnaire or by other household members.

The main imputation process starts with the replacement of item non-response on filter questions. The filter variable indicates whether or not a given individual holds a specific wealth component. If this information is missing, it is imputed by logistic regression, in each case using the specific set of covariates best suited for explaining variance in the missing filter information.⁵ Logistic regression is also applied for item non-response of the individual share of a specific wealth component because in the vast majority of cases, ownership is either 50%

⁴ Using the wealth data after imputation as described in this paper, the standard deviation of total net wealth is almost three times as large as in the previous version due to the mean-preserving nature of the imputation by Schäfer and Schupp (2006).

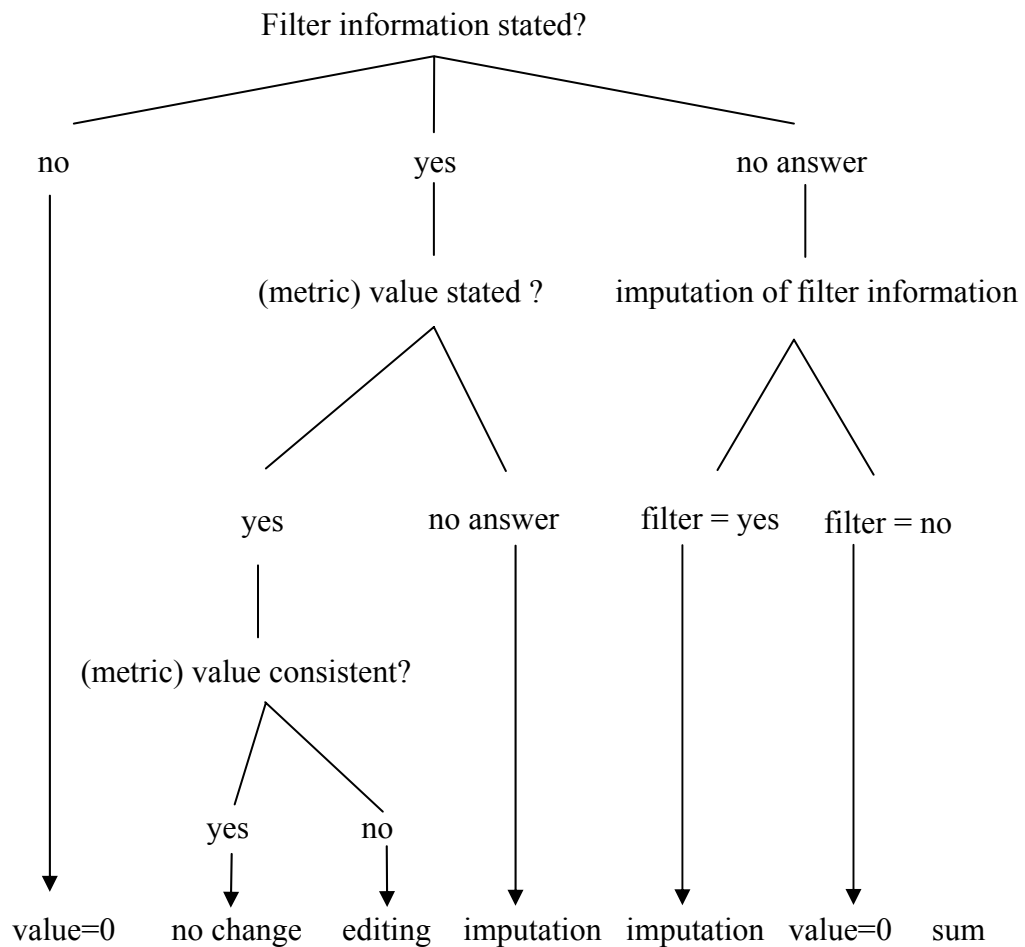
or 100%. The imputation of item non-response on all missing metric wealth data is based on Heckman selection models. These estimations are maximum-likelihood-based and consider clustering effects to control for sample selection (see Heckman 1979). This phenomenon is relevant for wealth questions included in SOEP if there is selection into ownership, i.e., if a respondent can only refuse to provide metric information on the value of a component conditional on being an owner of this component.

However, this procedure does not accurately mirror the uncertainty involved in the imputation process as such. This is why we reintroduce variance by adding a randomly chosen error term to the prediction based on the aforementioned regression model. Repeating this process five times, i.e., assigning five different error terms, yields a multiply imputed dataset.

A “decision tree” providing a brief overview of the general process of editing and imputation for each wealth component is presented in Figure 1. It also gives information on the absolute number of persons affected (and unweighted population shares). Note that the complexity of the entire process is considerably understated due to non-consideration of the imputation/editing of the individual share, the editing of the filter information and the imputation of partial unit non-respondents.

⁵ An exception is owner-occupied property, where the filter information was logically derived from the information on housing tenure in the household questionnaire or from information provided by other household members.

Figure 1: Simplified decision tree for imputing and editing wealth components



Component	n	%	n	%	n	%	n	%	n	%	n	%	n	%
own property	13704	57.4	7050	29.5	960	4.0	1102	4.6	507	2.1	569	2.4	23892	100
debt own property	13704	57.4	7371	30.9	541	2.3	1200	5.0	507	2.1	569	2.4	23892	100
other property	19923	83.4	2273	9.5	86	0.4	460	1.9	110	0.5	1040	4.4	23892	100
debt other property	19923	83.4	2492	10.4	44	0.2	283	1.2	110	0.5	1040	4.4	23892	100
financial assets	12294	51.5	8360	35.0	5	0.0	1822	7.6	587	2.5	824	3.4	23892	100
private pension	10948	45.8	8309	34.8	1	0.0	3308	13.8	689	2.9	637	2.7	23892	100
business assets	21473	89.9	833	3.5	0	0.0	350	1.5	64	0.3	1172	4.9	23892	100
tangible assets	20316	85.0	1618	6.8	0	0.0	592	2.5	53	0.2	1313	5.5	23892	100
consumer debts	20038	83.9	2244	9.4	0	0.0	366	1.5	53	0.2	1191	5.0	23892	100

3 Owner-occupied property

Perhaps the most important wealth component for private households in Germany is owner-occupied housing. In the SOEP, this component is surveyed as shown below, starting with a basic filter question followed (gross) market value, outstanding debts and the individual share of the property:

(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 %

The editing and imputation process for owner-occupied property can be divided into four steps. First, the answers given in the individual questionnaire are checked for consistency with the individual questionnaires of the other household members. Second, the individual answers are checked for consistency with information from the household questionnaire. Third, in selected cases, missing values are imputed using logical imputation. Fourth, a regression-based multiple imputation is conducted for remaining missing market values and total outstanding debt. The procedures for each of these four steps are described in more detail in the following.

3.1 Consistency check

The SOEP makes it possible to link information on housing tenure from the regular household questionnaire to wealth information on owner-occupied property from the individual questionnaire. Thus, one can obtain reliable information to conduct consistency checks on all of the other wealth components such as financial assets, which may still contain inconsistencies. Consistency becomes an issue particularly in cases of co-ownership within one household. If inconsistent, data on the filter, personal share, and both metric values (market value and debt)

may have to be edited. In the following we briefly describe the filter's intended purpose, the procedure used, and the number of individuals affected (in brackets).⁶

3.1.1 Filter of owner-occupied property⁷

Aim: To accurately identify the owner (or holder or proprietor) within the family. A first preliminary consistency check among all household members is conducted to clarify the ownership status, especially between parents and children. A property can be owned by parents, by children or by both parents and children.

- 1.) If both (adult) children and parents state that they are proprietors, and if the resulting sum of individual shares is greater than 100% and the market value of the property given by the individuals is of the same magnitude, the house is "assigned" either to the parents or to the children depending on the age structure of the persons involved (*20 cases*).
- 2.) If a child claims to be the sole proprietor and the parents also claim to be sole proprietors, and if the child gives a market value that differs strongly from the market value given by the parents, the filter for the child is set to "no owner". It is assumed that the child owns a second property or is in the process of moving to another property. The values given by the child are then assigned to the category "other property" (*5 cases*).

3.1.2 Market and debt value⁸

Aim: To obtain consistent information on the market value of the property and the outstanding debts of each owner in the household.

- 1.) If the values for "market values" and/or "debts" given by co-owners of the same property differ by not more than 30%, the average value is applied to the respective individuals (*812 cases for market value, 383 cases for outstanding debts*).
- 2.) Larger differences (measurement error) arise from one co-owner giving an exact amount in euros, and the other co-owner basically stating the same value but in thousands of euros, i.e., dropping the last three digits of the same amount as mentioned by the first co-owner. In

⁶ Certainly all those checks involve a certain degree of arbitrariness due to the normative nature of the plausibility controls.

⁷ The original SOEP variable name is SP85a01.

⁸ The original SOEP variable names are SP85a02, SP85a03 and SP85a04.

such (and similar) cases, the most plausible value is chosen on the basis of a case-wise check⁹ exploring regional information, size of housing unit, type of dwelling, etc. (*37 cases for market value, 6 cases for outstanding debts*).

3.) If the market value stated by one co-owner is twice that stated by the other co-owner, in most cases, the higher value is chosen as the “correct” market value after performing case-by-case checks. Here, the basic assumption is that the smaller value relates to the person’s individual share instead of the total market value of the property (*72 cases*).

4.) If the information provided by two co-owners (usually couples) on outstanding debts differs significantly, one of the two values is chosen and assigned to the other after performing case-by-case checks of occupancy, market value, income and monthly loan payments¹⁰. If neither of the two values seems more plausible than the other, the average value is taken (*77 cases*).

5.) If one co-owner states a positive value for the level of outstanding debts and his/her co-owner states “no debt”, the positive value is generally taken following case-by-case checks of other kinds of debts, monthly loan payments, occupancy etc. (*25 cases*).

3.1.3 Share of owner-occupied property¹¹

Aim: To prevent double-counting, that is, to ensure that the sum of the individual shares of one owner-occupied property does not exceed 100% within the same household.

1.) If both partners (or co-owners) claim to be sole owners (i.e., each owning 100%), or one partner claims to own 100% and the other states ownership of 50%, both shares are set to 50% (*394 cases*).

2.) If one person declares to be the sole owner and the other states that his/her share is x (with $0 < x < 50\%$), the first person’s share is set to $100 - x$ (*5 cases*).

⁹ A detailed list of every decision mechanism in the case-wise checks would be too complex for presentation in this paper, but can be provided on request.

¹⁰ Such checks consider potentially available longitudinal information on loan payments of the very same household in previous and subsequent waves.

¹¹ The original SOEP variable names are SP85a05 and SP85a06.

3.) If two persons in a household state the same share of more than 50%, it is assumed that this value gives the share both partners hold in common, and the remainder is owned by a third party not belonging to the household (*6 cases*).¹²

4.) If the overall household share is marginally less than 100% presumably due to rounding, the existing individual values are adjusted in order to achieve a sum of exactly 100% (e.g., 66% and 33% are changed to 67% and 33%, respectively) (*12 cases*).

3.1.4 Consistency with mortgage and interest payments from the household questionnaire

In more than 100 cases, households mentioned a mortgage on their dwelling in the household questionnaire (variables SH31 and SH32) while household members claimed in the wealth questionnaire that outstanding debt was zero. One possible problem with comparing this information from the household and individual questionnaires is that the household question on mortgage payments refers to the previous calendar year (2001), while the data on mortgage debts in the individual wealth questionnaire refers to the month of the interview (in 2002). One possibility here is that the household recently finished paying off its mortgage, in which case it would be logical that mortgage payments appear as zero in the individual questionnaire while debt information is given in the household questionnaire, and thus, no changes are required. Again, no changes are needed if payments were suspended for the reference period, if a new credit was taken out, or if the person who completed the household questionnaire did not contribute to mortgage payments. If the monthly mortgage and interest payments (SH32) were close to total outstanding debt (“close” being within a range of +/- 20%) the information on total debts was set to missing (and was thus integrated into the imputation process described below). In these cases, we assumed that the question had been misunderstood and that respondents had confused regular payments with outstanding mortgage debt (*55 cases*).

3.1.5 Implausible values

Market values for owner-occupied housing of less than 10,000 euros were checked on a case-by-case basis, controlling for size of housing unit, general condition and age of building, residential area, and outstanding debts (*17 cases*). Similarly, a case-by-case check for occu-

¹² This may be the case after a divorce from a former spouse now living outside the observed household but still holding a specific share of the property.

pancy, monthly loan payments and market value was done if the amount of outstanding debts was less than 2,000 euros (*2 cases*).¹³

3.2 Logical imputations

Before turning to the standard case of regression-based imputation of missing values (see Section 3.3 below), this section describes the logical imputation of market value for owner-occupied property and outstanding mortgage debt based on information given in the household questionnaire and other household members' individual questionnaires. We assume any valid information on owner-occupied housing given by other co-owners residing in the household affected by non-response to be superior to any other imputation routine, given that the information provided by (at least one of) the co-owners will consider the specific characteristics of the relevant property more explicitly than an imputation algorithm can do, the latter being subject to potential bias resulting from the restricted set of covariates (i.e., an omitted variable bias cannot be ruled out).

3.2.1 Imputation of missing filter information on owner-occupied property

Aim: To accurately define the proprietor within a family in case of INR and PUNR using information from the household questionnaire and information provided by other household members.

It should be noted that most of the cases dealt with in this section are affected by PUNR.

- 1.) Individuals with PUNR and those with INR on the central filter question asking for owner-occupied housing are set to “no owner” if the household is renting its home (*892 cases*).
- 2.) If 100% ownership of the dwelling is claimed by another party, PUNR and INR on the filter question are coded to “no owner” assuming that there exist no other potential co-owners (*419 cases*).
- 3.) If the sum of the co-owners' shares within the household is less than 100%, the filter of those individuals with PUNR/INR is imputed after case-wise checks of age and relationship to the head of household. It is assumed that very old and very young are not owners. If the filter is set to “owner”, the individual share for this individual takes on the value of 100%

¹³ In a few cases, the amount of outstanding debt exceeded the current market value. However, this was accepted based on the possibility that the home may have been purchased during the housing market boom (early-to-mid-1990s), but the price for the property came down considerably since then.

minus x , with x being the cumulative share of the remaining co-owners (*781 cases changed to “owner”, 183 cases changed to “no owner”*).

3.2.2 Imputation of missing market values and outstanding mortgage debt

Aim: To achieve consistent information among all (co-)owners in the household on market value and amount of outstanding debts.

- 1.) If a plausible value is stated by only one co-owner, this value is also assigned to the remaining co-owners (*137 cases for market value, 346 cases for outstanding mortgage debt*)¹⁴.
- 2.) If information on the level of outstanding mortgage debts is not given in the individual wealth questionnaire, information on monthly mortgage payments (variable SH32) from the household questionnaire is used to derive whether an individual is an outright owner (*523 cases changed to “debt-free”*). For the remaining 703 individuals with PUNR or INR on the target variables, the level of outstanding mortgage debt still needs to be imputed (see Section 3.3).
- 3.) For owner-occupying households with no valid information on monthly mortgage payments (variable SH32) from the household questionnaire, the following logical imputations are carried out: if the household has inherited the dwelling (*6 cases*) or if occupancy has lasted for more than 25 years, it is assumed that the household is debt-free (*12 cases*). In all other cases, the level of outstanding mortgage debt needs to be imputed (see Section 3.3.).

3.2.3 Imputation of missing personal shares of own property

Aim: To ensure that the sum of personal shares across all co-owners in a given household adds up to 100%. If the personal share of owned property is missing for at least one co-owner, we impute the missing information (as 100% minus the sum of all valid shares), i.e., we assume no ownership by parties living outside the household.

¹⁴ These numbers do not include changes for those with missing filter information.

- 1.) In cases where two co-owners fail to give information on their respective share (INR), both individuals are assigned 50% of ownership (*243 cases*).
- 2.) If one co-owner declares an individual share of x , the non-responding co-owner is assigned a share of 100% minus x (*112 cases*).
- 3.) Owners living in single households are declared to be sole owners (*34 cases*).

3.3 Regression-based multiple imputation of market value and level of outstanding mortgage debt for owner-occupied property

Aim: To obtain valid and consistent information for all owners regarding market value and outstanding mortgage debt on owner-occupied property.

The main idea of this approach is to conduct a regression of the variable of interest on the basis of all household members with valid information. The resulting estimates are assigned to otherwise comparable households with missing information on the dependent variable (hedonic regression). Thus, we reintroduce error terms in order to maintain variance and avoid regression-to-the-mean effects.

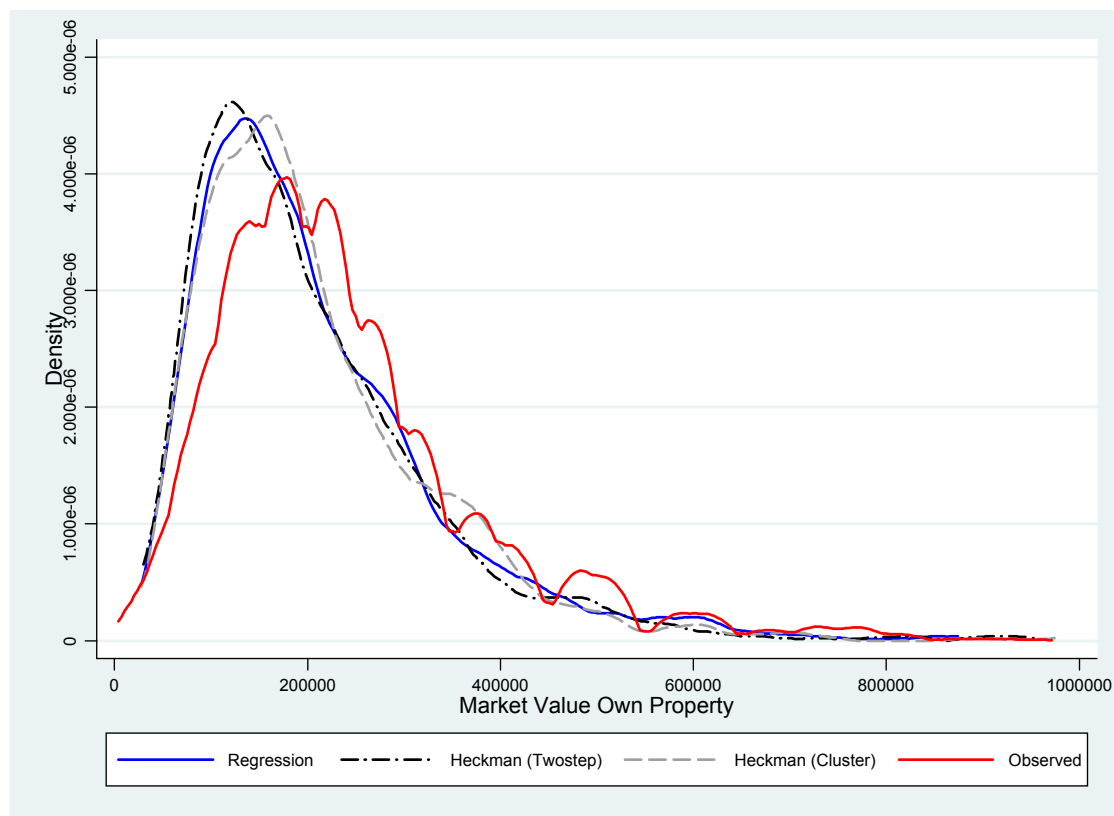
Given that information on market value and outstanding debt is identical for all co-owners in a given household, regression-based multiple imputation is carried out at the household level. One representative of every household is chosen by means of the following criteria: being one of the co-owners, having valid information on the variables of interest, holding the highest individual share of the property, having a relationship to the head of household. Not surprising, in many cases the selected household representative *is* the head of the household.

One exceptionally important decision is the selection of an appropriate estimation method to impute INR. We estimate a simple OLS regression and two versions of the Heckman selection model (Heckman's two-step estimator and Heckman's maximum-likelihood estimator using cluster information). A straightforward criterion to judge the applicability of a specific imputation method is to compare kernel density estimates of the predictions / imputations following the three different approaches with the "true" distribution. Figure 2 show results for the market value of owner-occupied property. The **red** line represents the information observed, the **blue** line displays the density of imputed values based on an OLS regression, the **black** line shows the results following the Heckman's maximum likelihood estimator control-

ling for clustering effects, and finally, the **grey** line gives the density of an imputation based on Heckman’s two-step estimator.

Although the results of the three different approaches do not differ much, we opt for the maximum-likelihood based Heckman selection regression model. We do so in order to control for sample selection – which is not possible by means of OLS regressions – as well as to control for regional clustering effects inherent in the sampling design of the SOEP sample.

Figure 2: Comparison of different imputation methods to the observed cases of market value of own property¹⁵



As market values and outstanding debt are strongly correlated, they serve as mutually independent variables if one is missing. First, households with an unknown level of debt are assigned the average of all households with valid positive debt information, i.e., larger than zero. Accordingly, all the observed households in the regression for the market value have valid information on debt. The regression-based¹⁶ predictions of the market values for households affected by INR/PUNR, together with observed market values, are used as covariates for the next step in the imputation process: the assignment of outstanding mortgage debt. In

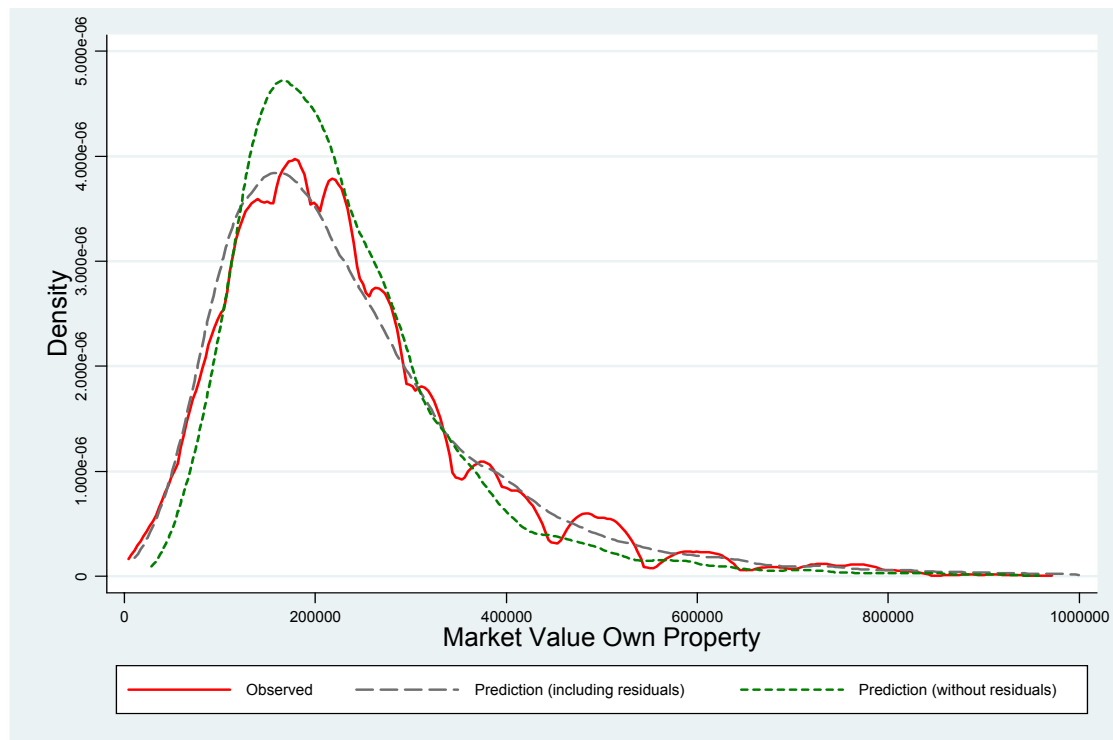
¹⁵ All three imputed values include randomly drawn residuals. The cut-off point for this graph was 1,000,000 euros. The analyses are unweighted and carried out at household level.

¹⁶ See Table 13 in the appendix for the covariates used in each regression.

an iterative process, the predicted debt values are used again for a second prediction of market values, and so on for further predictions. This process is repeated four times (R^2 in the fourth iteration is 0.52 for market values and 0.59 for outstanding mortgage debt). After the fifth regression of the market value, we add a randomly chosen residual (from the true distribution with non-missing values) to the predicted value for the missing observation. This process is thought to maintain variance.¹⁷

Focusing only the population that provided data on the market value of their home, Figure 3 compares the kernel density estimates of the distribution derived from the predicted values (based on the Heckman selection model controlling for clustering effects) with the observed information. The **grey dashed** curve gives the density of the imputed values *with* residuals and the **green dotted** curve gives the density *without* randomly chosen residuals in contrast to the observed market values (**red** curve). Apparently the grey dashed and the red curve match very well, although not considering residuals does yield a distinct regression-to-the-mean phenomenon. That is, the variance will be greatly underestimated if no residuals are considered in the imputation process.

Figure 3: Predicted market values with and without randomly chosen residuals¹⁸



¹⁷ The distribution of the residuals was trimmed at the 0.5 and the 99.5 percentiles in order to reduce the impact of extreme outliers.

Nevertheless, there is uncertainty imbedded in the imputation process itself. This problem can be dealt with best by repeating the imputation k times, i.e., by doing a multiple imputation. We use $k = 7$; that is, the process of randomly drawing a residual is performed seven times to obtain seven different imputations (=implicates) for the market value (and all other missing metric values described below). These imputed market values are again used for seven different regressions of the outstanding level of mortgage debt. Of these seven, for every household the highest and lowest values are deleted. This is done to ensure a reasonable match of market value and outstanding mortgage debt assuming that the debt – in principle – should not be considerably higher than the market value.¹⁹

Finally, given that the regression is performed at the household level, the imputed values are assigned to all other co-owners within the same household, yielding five (gross) market values and five values for the outstanding mortgage debt (variables P0100x, where $x = a, b, \dots, e$ identifies the five implicates; see Section 10 on variable naming conventions).

Using kernel density estimates, Figure 4 compares the distributions of the truly observed cases (**red** curve) with the five implicates based on predictions of this very same population (5 **green** curves) and the five implicates based on predictions of the population with missing data due to INR/PUNR. We find only slight variation resulting from the in-sample prediction, i.e., within the group of fictitious implicates for the observed cases.²⁰ Much more variation, however, results from the five implicates for the non-responding cases, i.e., the out-of-sample predictions. Most important, comparing the distribution of the truly observed cases with that of the imputed non-responding cases, we find clear indications that non-response is not random: obviously there is a concentration in the lower part of the wealth distribution that is not driven by the imputation procedure being downward-biased. Instead, as shown by more detailed analyses, lower market values often result from apartments being in older buildings, in rural areas, smaller in size, or occupied by elderly persons with a long occupancy period

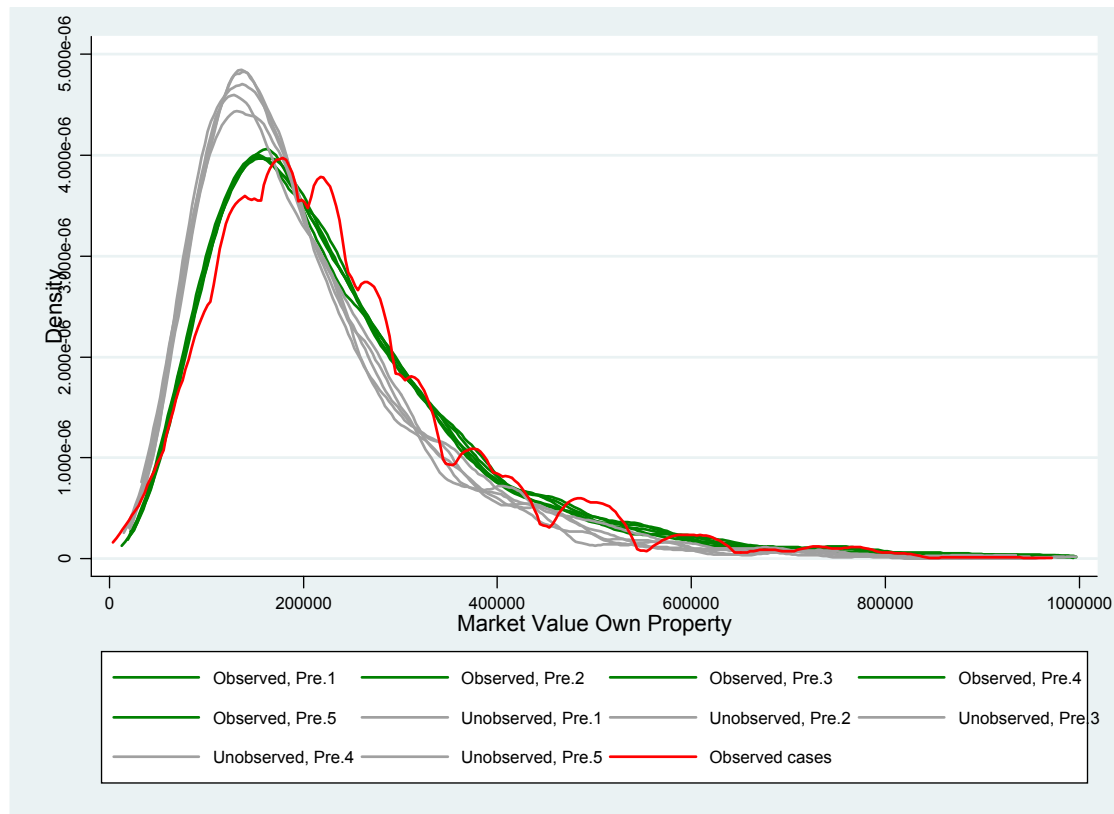
¹⁸ Values higher than 1 million euros are trimmed in this figure. Displayed are the values of all households with an observed market value (5104 households).

¹⁹ However, this happens even if the highest and lowest values are deleted. For the remaining cases where the debts exceed the market value by more than 1.5 times, new residuals for the level of debts are drawn. This was done for less than a dozen households per variant.

²⁰ In fact, these five green lines represent variations of the grey dashed curve in Figure 3.

(which again may be a reason for lacking knowledge of the current housing market and inability to accurately estimate the market price).

Figure 4: Comparison the distribution of implicates for cases with non-response (“unobserved predictions”) and for implicates for observed cases (“observed predictions”) with the distribution of truly observed cases²¹



In order to support methodological research on such issues as the impact of imputation on wealth inequality, all wealth variables generated are assigned a corresponding flag variable. These flag variables take on a value of “1” if the respective information was edited, a value of “2” if imputation took place and “0” otherwise. A more detailed description of the variables generated is given in Section 10 below.


For welfare-oriented analyses, a *net* value of owner-occupied housing might be more appropriate. This *net* value bound in owner-occupied property for an individual is given by the difference between gross market value (variable P0100x with x=a, b, ..., e for the five implicates) and the outstanding mortgage debt (P0010x) multiplied by the person’s individual share (P00010).


²¹ Values higher than € 1,000,000 are trimmed in this figure. Implicates include randomly chosen residuals.

4 Other property

The second wealth component is on “other property”²². This set of variables encompasses the following information: the corresponding filter variable (SP85b01), the type of property (SP85b02 – SP85b06), the number of other properties (SP85b07), the gross market value of thereof (SP85b08), the personal share (SP85b09/SP85b10), and the sum of outstanding debts related to this other property (SP85b11/SP85b12).

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

²² Other property here refers to real estate that is owned by a given individual but not used as the principal residence.


4.1 Internal checks of consistency and logical imputations

Aim: To achieve consistent wealth information in case of co-ownership of “other property”.

See Table 1 for a description of how values of “other properties” are edited and imputed.²³ Given the lack of information about other properties in the household questionnaire (which could have been used as an “external” benchmark) as well as the potential heterogeneity of the components included in this category, the philosophy for changing data is a rather conservative one. In other words, data is only edited/imputed if the basic information provided by co-owners living in the same household (mostly couples) with respect to type and amount of other property is not contradictory.²⁴ In total, 177 cases were changed in this process.

Table 1: Editing and imputation process for “other property” (co-ownership households only)

Legend

d_1, d_2 :	level of debt (SP85b11/SP85b12) for partner 1 and partner 2
m_1, m_2 :	market value (SP85b08) for partner 1 and partner 2
s_1, s_2 :	personal share (SP85b09/SP85b10) for partner 1 and partner 2
$d_i < 0, m_i < 0, s_i < 0$	respective information is missing for person i (INR, UNR)
–	no observations
/	no editing/imputation
	regression-based imputation for one or more values
“check”	checking for measurement error due to missing digits (i.e., value given in 1,000 Euros) or sum of individual shares exceeding 100%
bold	editing
<i>italics</i>	logical imputation
(n)	number of cases changed

²³ The table is to be read as follows: Cell A2 means the head of household and his/her partner state the same value for debts and both partners own 50% of the property each. Only the market values they stated are slightly different (deviation by less than one-third). In this case, the mean of the two market values was calculated and ascribed to both partners. Bold type indicates that the values have been edited and the number in brackets states that this procedure was carried out for 42 cases.

²⁴ It is also checked whether the market value stated is too low due to missing digits, mostly zeros [e.g., 200 instead of 200,000 euros] (4 values are edited to a different value, and 2 values are set to missing with consequential need for further imputation). There are two more changes which are not displayed in Table 1: In one case, the head of household stated a share of 100% and his partner a share of 25%, although they gave the same market value, the same level of debt and the same amount and type of property. In this case, the ownership-ratio is changed to 75:25 (one case). In another case, two partners indicated the same amount and type of other properties and different shares (which amounted to 100% in total), but only one partner gave information about market value and level of debt. Here, the latter partner’s information was used to substitute the former’s missing information (one case).

If in the household questionnaire data on income from property (including income from renting and leasing) was given (variable SH41), but no information on other properties was given in the individual wealth questionnaire for the very same household, no changes were made. This may happen on those few occasions where property had been sold since the previous year, given that the household information is surveyed for the previous calendar year.

			A	B	C	D	E
Debt	Share		50 : 50	50 : s ₁ , s ₁ <0	s ₁ <0, s ₂ <0	100 : 50	100 : s ₁ , s ₁ <0
	Value	No					
d ₁ =d ₂ , d ₁ >0, d ₂ >0	m ₁ =m ₂ , m ₁ >0, m ₂ >0	1	/	s ₁ =50 (3)	s ₁ =s ₂ =50 (4)	s ₁ =s ₂ =50 (5)	s ₁ =s ₂ =50 (14)
	m ₁ <m ₂ +1/3m ₁ , m ₁ >m ₂ >0	2	m ₁ =m ₂ =(m ₁ +m ₂)/2 (42)	-	-	/	m ₁ =m ₂ =(m ₁ +m ₂)/2 s ₁ =s ₂ =50, (2)
	m ₁ >m ₂ >0, m ₁ >m ₂ +1/3m ₁	3	/	-	-	/	-
	m ₁ >0, m ₂ <0	4	m ₁ =m ₂ (13)	-	-	/	-
	m ₁ <0, m ₂ <0	5	/	s ₁ =50 (1)	s ₁ =s ₂ =50 (2)	/	-
d ₁ <d ₂ +1/3d ₁ , d ₁ >d ₂ >0	m ₁ =m ₂ , m ₁ >0, m ₂ >0	6	d ₁ =d ₂ =(d ₁ +d ₂)/2 (10)	-	-	-	-
	m ₁ <m ₂ +1/3m ₁ , m ₁ >m ₂ >0	7	d ₁ =d ₂ =(d ₁ +d ₂)/2 m ₁ =m ₂ =(m ₁ +m ₂)/2 (22)	-	-	-	-
	m ₁ >m ₂ >0, m ₁ >m ₂ +1/3m ₁	8	check (2)	-	-	/	-
	m ₁ >0, m ₂ <0	9	-	-	-	-	-
	m ₁ <0, m ₂ <0	10	-	-	-	-	-
d ₁ >d ₂ >0, d ₁ >d ₂ +1/3	m ₁ =m ₂ , m ₁ >0, m ₂ >0	11	check (11)	-	-	-	-
	m ₁ <m ₂ +1/3m ₁ , m ₁ >m ₂ >0	12	/	-	-	/	-
	m ₁ >m ₂ >0, m ₁ >m ₂ +1/3m ₁	13	/	-	-	-	-
	m ₁ >0, m ₂ <0	14	check (3)	-	-	/	-
	m ₁ <0, m ₂ <0	15	/	-	-	-	-
d ₁ >0, d ₂ <0	m ₁ =m ₂ , m ₁ >0, m ₂ >0	16	d ₁ =d ₂ (6)	s ₁ =50 d ₁ =d ₂ (1)	-	-	-
	m ₁ <m ₂ +1/3m ₁ , m ₁ >m ₂ >0	17	m ₁ =m ₂ =(m ₁ +m ₂)/2 d ₁ =d ₂ (6)	m ₁ =m ₂ =(m ₁ +m ₂)/2 s ₁ =50, d ₁ =d ₂ (2)	-	/	-
	m ₁ >m ₂ >0, m ₁ >m ₂ +1/3m ₁	18	check (2)	check (2)	-	-	-
	m ₁ >0, m ₂ <0	19	d ₁ =d ₂ m ₁ =m ₂ (8)	s ₁ =50, d ₁ =d ₂ m ₁ =m ₂ (4)	-	-	/
	m ₁ <0, m ₂ <0	20	d ₁ =d ₂ (3)	s ₁ =50, d ₁ =d ₂ (1)	-	-	/
d ₁ <0, d ₂ <0	m ₁ =m ₂ , m ₁ >0, m ₂ >0	21	/	-	-	-	s ₁ =s ₂ =50 (2)
	m ₁ <m ₂ +1/3m ₁ , m ₁ >m ₂ >0	22	-	-	-	-	-
	m ₁ >m ₂ >0, m ₁ >m ₂ +1/3m ₁	23	-	-	-	-	-
	m ₁ >0, m ₂ <0	24	m ₁ =m ₂ (4)	-	-	-	/
	m ₁ <0, m ₂ <0	25	/	s ₁ =50 (2)	/	-	/

4.2 Imputation

Any information on the filter variable, the personal share, the market value and level of outstanding debts still missing after the procedures described in Section 4.1 are imputed as follows. In contrast to Section 4.1 however, the following procedures are applied not only to co-owning households but to all individuals therein, irrespective of household structure.

4.2.1 Imputation of the filter variable

Two different logistic regression models are used to impute missing filter information: one model for cases with INR and one for cases with PUNR. For both models, we use the same set of covariates except for variables derived from the individual questionnaire in case of PUNR (information on sex and age is available from the household grid for all household members irrespective of their response behavior). All predicted values of less than 0.5 are rounded down to zero, which means it is assumed that the person does not possess this wealth component. If the predicted value is greater or equal to 0.5 the filter information is set to 1, indicating that this individual owns other property. Covariates used in this regression are described in Table 13 in the appendix.

4.2.2 Imputation of the personal share

For those who own other property but do not state their personal share, this information is imputed based on an OLS regression (see Table 13 in appendix for more details).

4.2.3 Imputation of market value and debts

Given that there is no filter for debt on other property, if relevant debt information is lacking, the existence of debts has to be imputed (see variables SP85b11/SP85b12). If the household questionnaire states that no payments are being made on other property (variable SH4202=0), debt is assumed to be zero. However, if payments *are* being made an unknown level of debt is assumed, and this has to be imputed.²⁵

If information on other property is not available from the household questionnaire (e.g. in case of INR) but the filter information indicates the presence of other property in the individual questionnaire, a logistic regression is applied to determine if the individual has debts on

²⁵ If the household questionnaire indicates other properties (variable SH40=1) but the question on mortgage repayments has been refused, we also assume positive debts with consequential need for imputation.

other property. Predicted values below 0.5 are rounded down to zero, that is, it is assumed that the individual is debt-free. Predicted values greater than or equal to 0.5 indicate existing debts which will have to be imputed.

The actual imputation of missing data on market value and on level of outstanding debt for other property follows the same logic as imputation for owner-occupied housing (see Chapter 3.3 above). In an iterative process, values for other property are imputed using a maximum-likelihood-based Heckman selection model controlling for clustering effects. Finally, in order to maintain variance in the resulting estimate and to consider the uncertainty in the imputation process, we repeat the process of adding randomly chosen error terms five times, yielding five different imputates per individual for gross market value and eventual debt associated with other property.²⁶

5 Financial assets

Financial assets are the most important category of wealth in Germany, but given the large degree of heterogeneity in the potential components thereof one can assume higher non-response here than in the case of owner-occupied property, which mostly refers to a single object only. Given this and keeping in mind the difficulty of achieving a high response rate when collecting information on such complex issues, it was decided to ask only those individuals with a “significant” amount of wealth for information on their financial assets, setting the threshold at 2,500 euros. The information gathered on this wealth component is the filter variable (SP85c01), the total value of the assets (SP85c02) and the personal share of the assets held by the individual (SP85c03/SP85c04).

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 %

²⁶ In case of INR on the number of other property components (n=199) this value is set to the mode of the true distribution, which is 2. This assumption is made due to otherwise lacking appropriate information for imputation.

5.1 Checks of consistency

Only a few individuals answered the filter question with “no” but then gave further information on holdings in the follow-up questions. In such cases, the latter values are set to “not applicable / -2” (4 cases).

Other consistency checks using information on financial assets from the household questionnaire (variables SH4301, SH4304, SH4305) are not applied. Especially in households that are less well-off financially, the problem arises that although none of the individual household members owns above the 2,500 euro threshold used in the individual wealth questionnaire, it cannot be ruled out that overall financial assets aggregated across all household members exceed this amount.

5.2 Logical imputations for co-owners

Aim: To ensure consistent information among co-owners. Changes are performed only if there is a strong indication that the head of household and spouse/partner share their financial assets equally (i.e., 50% each). After logical imputation, the value of financial assets is identical for all individuals who appear to own the same (set of) financial assets. Thus logical imputation is only conducted in very few cases (see Table 2).

Table 2: Logical imputation/editing of financial assets

Share \ Value	$a_1 = a_2 > 0$	$a_1 > 0, a_2 < 0$ (in CAPI-based interviews the eventually available categorical value for a_2 must be consistent with the metric wealth information given by person a_1)
$s_1 = s_2 = 50$	/	$a_1 = a_2$ (62 changes)
$s_1 = 50, s_2 < 0$	$s_2 = 50$ (30 changes)	$s_2 = 50, a_1 = a_2$ (22 changes)

Legend

$a_1, a_2:$	financial assets (SP85c02) of partner 1 and partner 2, respectively
$s_1, s_2:$	personal share (SP85c04) of partner 1 and partner 2, respectively
$s_i < 0, a_i < 0$	respective information is missing for person i (INR, UNR)
/	no editing/imputation
(n)	number of cases changed

5.3 Imputation

In this step, any remaining missing information on either the filter question or the personal share or the market value is imputed using a list of independent controls which vary slightly across the models (see Table 13 in the appendix for more details).

5.3.1 Imputation of the filter variable

Aim: To state for all individuals whether the person owns financial assets worth at least 2,500 euros.

The imputation of the filter variable is based on two logistic regression models: one for PUNR and one for INR.²⁷ Predicted values below 0.5 are set to zero if these persons are assumed not to own the particular wealth component. If the predicted value is 0.5 or higher, the filter information is set to 1 assuming the individual holds financial assets worth at least 2,500 euros.

5.3.2 Imputation of the personal share

For those who own financial assets but do not state their personal share, a value is imputed based on logistic regression. Here, the dependent variable is either a 50% share or sole ownership (i.e., 100% share), given that more than 98% of the observed cases state that their personal share is either 50% or 100%. All predicted values of less than 0.5 are rounded to zero, assuming a personal share of 50%. All higher values are rounded to one assuming sole ownership of the financial assets stated.

5.3.3 Imputation of the value of the financial assets



A maximum-likelihood-based Heckman selection model controlling for clustering effects and randomly chosen error terms is applied to derive five different imputates; this procedure was already described in Section 3.3 above).²⁸

²⁷ The respective Pseudo R² of those regression estimates is 0.42 for the model focusing on INR, and 0.40 for those cases with PUNR.

²⁸ The 0.5% smallest and the 0.5% largest residuals are excluded from this process.



6 Private pension assets, business assets, tangible assets, debt from consumer credits

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 forementioned?

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!

The editing and imputation process used for data on private pensions, business assets, tangible assets and debt from consumer credits (variables SP85d01-SP85g02) is described below in a single section given that they all follow a similar logic requiring similar treatment of inconsistent or missing values.

6.1 Checks of consistency/logical imputations

There is no sufficiently comparable (metric) information available in the household questionnaire on any of those components. Standard SOEP data includes qualitative information on the existence of various kinds of assets and the total amount of interest and dividends received from these investments. However, none of these components correspond perfectly with the wealth components defined in the 2002 individual questionnaire.

It should also be noted that there is always a chance that small individual wealth holdings do exist below the threshold of 2,500 euros (i.e., “no” wealth), particularly when qualitative information at the household level suggests the existence of such wealth.

6.2 Imputation

Missing information on the filter, the market value and – if applicable – the personal share of the four components considered here is imputed as follows:

6.2.1 Imputation of the filter variable

Missing filter data is imputed by means of logistic regression. For each component, there are separate models for INR and PUNR, each using individual information on sex and age as well as a wide range of covariates from the household level. The exact list of control variables, however, slightly differs for the various wealth components (see Table 13). In line with the procedures described above, predicted values below 0.5 are rounded down to zero, assuming that the person does not own the respective wealth component and vice versa for predicted values greater or equal to 0.5.

6.2.2 Imputation of the personal share

For the four components considered here, respondents are asked to specify their personal share only in the case of business assets. Missing personal shares are imputed using a logistic regression model estimating the probability that someone is sole owner or co-owner of an enterprise, the latter being interpreted as a personal share of 50%.

6.2.3 Imputation of the value

A maximum-likelihood-based Heckman selection model controlling for eventual regional clustering effects arising from the sample design is estimated (see also Section 3.3) using

covariates as described in Table 13. Adding five randomly chosen error terms to the value resulting from the out-of-sample prediction again yields a set of five different implicates.

7 Restrictions of the 2002 SOEP wealth information

With respect to the overall coverage of wealth held by the German population in 2002, the aggregated wealth measure is restricted by the lack of information on wealth holdings by children due to the fact that household members are first given an individual questionnaire the year they turn 17. Although this design ignores children's wealth holdings, this should be of minor relevance given that children do not generally possess large amounts of wealth. Another, potentially more relevant shortcoming (depending on one's definition of wealth) arises from the exclusion of cars when measuring tangible assets²⁹.

A further possible concern is the introduction of a lower threshold value of 2,500 euros for three of the surveyed (wealth and debt) components. While this intends to reduce the response burden on interviewees by ignoring data of negligible amounts³⁰, it may yield a systematic underestimation of total wealth aggregates (see Section 8 below). It should be noted that in the replication of this wealth module for 2007, such lower thresholds were completely abolished. This data will be used to analyze eventual selectivity arising from these thresholds and will make it possible to quantify the degree of underestimation in the overall share of wealth holders as well as in the overall stock of wealth.

8 Comparison with the national aggregate information

The empirical coverage of aggregate household wealth in the 2002 SOEP survey can be compared to information from national aggregated statistics (such as national accounts or *Bundesbank* data). Such micro-macro comparisons are often used as an indication for the quality of the micro data at hand (see Table 3). However, whenever making such comparisons one should keep in mind that the precise definitions of the aggregates almost always differ in comprehensiveness. Thus, it is important to know which components are actually included in a given aggregate and whether it is measured in gross terms or in net terms, taking into ac-

²⁹ SOEP does not collect information on durables or personal belongings in a household except for tangible assets in the form of gold, jewelry, coins or other valuable collections.

³⁰ However, if an individual holds several wealth components below the threshold, this may cumulate to a considerable underestimation of the true total wealth holdings.

count any taxes or possible costs involved. Obviously, omissions or exclusions—such as that of private cars from the category of tangible assets in SOEP—will result in significant differences from national aggregates that include the value of these items.

Table 3: Comparison of total wealth of private households with the 2002 national balance sheet (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: Excluding non-profit institutions, excluding currency and transferable deposits, certain claims on insurance corporations (e.g. 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: Bartzsch & Stöss (2006), Table 10: Financial assets and debt of German households (without entrepreneurial loans).

5: Mainly for commercial and consumption purposes.

6: The national balance sheet includes all personal belongings (Ammermüller et al. 2005, p. 100).

Source: Ammermüller et al. (2005: 100ff); Bartzsch & Stöss (2006); SOEP; authors' calculations.

The largest wealth component in Germany is real estate, consisting of owner-occupied property and other property. Making up 98% of the corresponding national aggregate, this type of wealth appears to be only marginally understated in the SOEP data. Things look rather different for financial assets, where we find a strong underestimation with only 34% of the national reference value being covered by the SOEP. Some important factors contribute to this apparent difference. First, certain items are fully included only in the financial accounts data, such as “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 in all, these items add up to about € 950 billion” (Bartzsch and Stöss 2006, p. 12). These items are not fully included in the SOEP survey mainly because respondents often

simply cannot provide this information. This is especially true for insurance claims. Another important factor is the inclusion of non-profit institutions in the national balance sheet wealth data, whereas SOEP surveys only private households. This amounts to another € 150 billion. If these components are excluded from the national accounts reference value, SOEP's coverage of financial assets rises to almost 50%. The coverage of liabilities in the SOEP is much higher (about 93% of the macro comparison value), with mortgage debt being better represented than other debts.

Another prominent discrepancy results from the accounting of durables. While the national balance sheet data includes all personal belongings, SOEP surveys only information on tangible assets in the form of gold, jewellery, coins and any valuable collections. Consequently, SOEP coverage amounts to less than 10% of the clearly more comprehensively defined macro information.

All this can be interpreted to mean that the “big numbers” such as owner-occupied housing and associated liabilities can be captured well in a rather simply structured wealth module such as the one employed in SOEP 2002. At the same time, the more heterogeneous and diversified wealth holdings are (e.g., in case of financial assets), the more problems of coverage arise. Excluding the category “durables and tangible assets” and considering the adjustments for financial assets (i.e., using the definition “Financial Assets II” in Table 3), the aggregated gross total wealth in SOEP is about 82% of the corresponding macro information (i.e., 6,493 billion euros / 7,925 billion euros). Considering the high coverage of liabilities, the comparison for total net wealth yields roughly 80%, which is an extraordinarily good result from a cross-national perspective (see results for other various country data sets available in the Luxembourg Wealth Study in Sierminska, Smeeding and Brandolini 2006).

9 Impact of imputation on wealth inequality

Tables 5-8 reflect the variation across the five versions resulting from the multiple imputation process. Using adequate cross-sectional weighting factors correcting for eventual differences in sampling design and selective response behavior, these tables present basic statistics (mean, median, standard deviation, selected quantiles, minimum, maximum) and inequality indices (Gini coefficient, mean log deviation (MLD), half squared coefficient of variation (HSCV)) for the five implicates of each wealth component. In the last column, the mean across all five versions is displayed. The lowest value of each statistic is marked in green and bold; the high-

est value in red. Table 5 shows results for all individuals successfully interviewed (i.e., excluding PUNR), while Table 6 shows results only for those actually holding the particular (wealth or debt) component. Table 7 and 8, respectively, present the same statistics at the household level, allowing for the incorporation of wealth holdings by partial unit-non respondents (PUNR).

As expected, differences across the five implicates based on the total population (see Table 5) are only marginal. To give an example, the mean of the market value of owner-occupied property varies between 50,674 and 50,891 euros (about 0.4% of variation). However, while this is true at the aggregate level this must not be (and in fact is not) the case for the five implicates of a given individual with INR. Due to the incorporation of randomly chosen error terms, these five values can vary considerably.

A comparison of the results based on the imputation procedures described here (Frick, Grabka and Marcus, hereafter FGM2007) with those derived by Schupp and Schaefer (SS2006) using an earlier and less sophisticated imputation approach is presented in Table 9. A large difference can be seen in the number of observations which underwent editing and imputation. The techniques used by FGM2007 obviously affect a much larger number of observations resulting from the additional imputation of PUNR and the extensive checks for consistency—e.g., the market value for owner-occupied property is “treated” for about 1,100 individuals by SS2006, while this is true for almost three times as many observations in the more comprehensive approach described in the paper at hand.

While this was to be expected in principle, it is interesting to see the impact of this differential treatment through editing and imputation on average values, aggregates and inequality. It should be noted that even the mean values for “observed” cases might differ between SS2006 and FGM2007 given that the latter also edit observations identified as inconsistent, while these cases were not affected by the treatment used in SS2006. In fact, for those wealth components in which the numbers of “observed” cases are identical in Table 9, we also find identical means for the two approaches. Due to the increased number of observations with positive wealth holdings, the corresponding aggregated figure for each wealth component increases significantly. According to the FGM2007 approach, approximately 25% of net wealth was imputed (see Tables 10 and 11 for results obtained on the basis of individual and household level, respectively). For the various components, this share ranges from less than 20% to almost 50% (in case of business assets).

It is not ex ante clear how mean and median values might change using the different imputation strategies. The FGM2007 version, with the exception of the very heterogeneous components business assets and debt from other property, results in slightly lower mean values while the mean-preserving imputation method applied by SS2006 yields rather similar values for the observed and the imputed observations. Applying Heckman selection regression models in combination with an assignment of randomly chosen error terms in the updated imputation process, on the other hand, not only takes into account the potential selectivity built into the missing process but also ensures better preservation of the variance. As such it can be expected that variation and inequality measures differ considerably between data resulting from the two imputation strategies.

When looking at the impact of imputation on wealth distribution for the entire population, imputation has an inequality-reducing effect (although this reduction is not always statistically significant) (Table 12). The HSCV is always considerably more reactive to imputation than the Gini coefficient. This can be explained by the very nature of the imputation procedures: in most cases, observations treated by means of editing and imputation do have positive wealth holdings. As such, these values often are important at the upper tail of the distribution and simply increasing the number of “rich” observations effectively reduces inequality in the upper end of the distribution. This result would be reversed if most of the newly added observations were even richer than the richest observations prior to imputation. Repeating this exercise only for observations with positive values on the respective wealth and debt components, we find that although inequality is generally reduced due to imputation and editing, there are some ambiguities. For owner-occupied property, financial assets, pension assets and debts related to owner-occupied property, the effect of imputation on inequality is now smaller, whereas it is larger for other property (and related debts) as well as for business assets. However, this effect is insignificant in most cases.

10 Conventions on naming variables

The generated SOEP wealth data is stored in two separate data files called PWEALTH for information at the individual level and HWEALTH for correspondingly aggregated data at the household level. These data refer to the survey year 2002 and may be supplemented by similar information from future waves. The first repetition of the wealth module will be carried out in the survey year 2007 and will have to undergo a similar treatment with respect to imputation and editing before being stored in these intentionally cumulative files. Consequently, the variable SVYYEAR (survey year)³¹ will be a constant (2002) in this first release.

Wealth-related variable names in the file PWEALTH consist of six digits. The first digit tells the user which wealth component is referred to, and the second to sixth digits provide more detailed information about possible filter information, the personal share, the gross amount, and the amount of any outstanding debt. In principle a digit is coded “1” if a given variable does indeed contain this specific piece of information and “0” otherwise. The code “2” indicates that this is the flag variable, showing whether or not the corresponding wealth information was imputed or edited.³²

To give an example, the variable P0111A represents the net value of owner-occupied property considering the individual share an individual holds. The first digit (P) indicates the wealth component (here real property of primary residence). The second digit identifies whether the information is related to a filter question. Here it is set to zero because this variable is not related to a filter question, but rather, as shown by the third digit, to the market value of a given wealth component. The fourth digit informs the user that this variable also considers the amount of debt needed to derive a *net* wealth measure. The fifth digit indicates that the individual share is taken into account by the variable. Finally, the sixth digit identifies the five imputates obtained from the multiple imputation procedure ranging from “A” through “E”.

Variables at the household level in file HWEALTH are generated in a similar way. In this case, however, the fifth digit is always set to “H”, indicating that these variables are aggre-

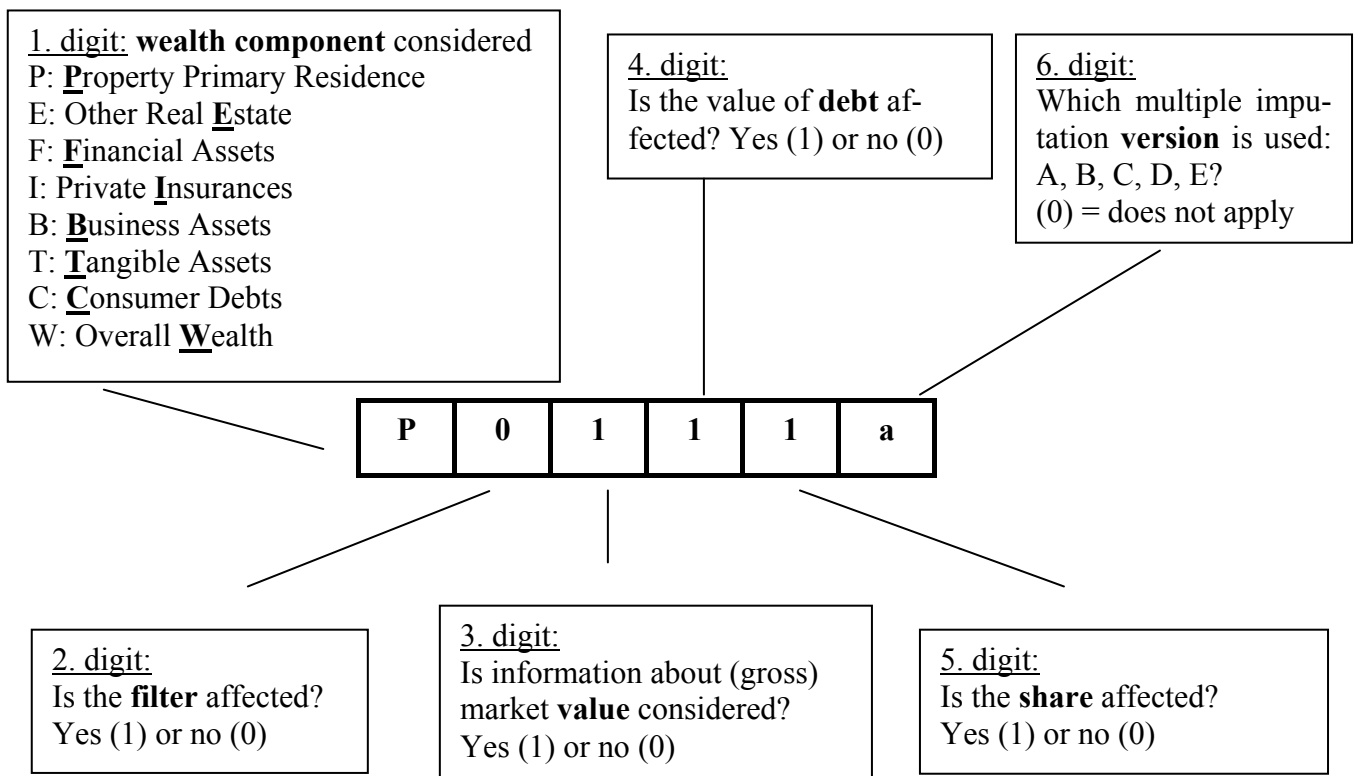
³¹ In other SOEP data files, the functionally equivalent information may be named ERHEBJ (e.g. in the spell data).

³² These flag variables carry a value of “2” for imputed observations, a value of “1” for those which were edited, and are coded “0” otherwise.

gated at the household level. Eventually, this aggregate includes wealth information imputed in case of PUNR.

The wealth information in the 2002 SOEP questionnaire is surveyed at the individual level and thus also imputed or edited at the individual level (although checked against household information for consistency). The three different aggregation levels that result are relevant in cases of co-ownership.

Figure 5: Variable naming conventions (file PWEALTH)



The first aggregation level is the individual level. It reports information on the share of a given wealth component the individual actually possesses. To obtain this “individual” information, a given market value referring to the object (e.g., house) needs to be multiplied by the individual percentage share operationalized by a value between zero and 100 in case of sole ownership. To give an example: the individual share of the market value of financial assets (variable F0101x with “x” referring to any of the five implicates) results from the multiplication of the market value of financial assets (variable F0100x) by the individual percentage share (F00010) (see Table 4 below).

Table 4: Illustration of the different aggregation levels (using information on Financial Assets)

	1. aggregation level = individual			2. aggregation level = couple	3. aggregation level = household
HHNRAKT	PERSNR	F0101A “Personal Share of Mar- ket value”	F00010 “Personal Share (%)”	F0100A “Market Value”	F010HA “Market Value”
	available in file PWEALTH				available in file HWEALTH
2211	22101	12500	50	25000	32641.47
2211	22102	12500	50	25000	32641.47
2211	22104	7641.47	100	7641.47	32641.47

The second aggregation level is the couple (legally married or cohabitating). Here, the information on the aggregate for a given wealth component held by a couple is reported. To give an example: the market value of financial assets (variable F0100x) reports all financial assets shared by the couple.

Finally, the third aggregation level is the household. Here, the amount of the total value of a given wealth component for all household members is given. To give an example: the household-level market value of financial assets (variable F010Hx) is the sum of all individual shares of financial assets (F0101x) across all household members. As such, in multi-person households with several co-owners, there is double counting in all variables carrying information on the market value of a given wealth component. Due to the additional consideration of the information on the share actually owned by an individual, there is no double counting problem in the household-level data.

10.1 Variable list at the individual level

Identifiers

PERSNR	Individual identifier
HHNRAKT	Wave specific household identifier
SVYYEAR	Survey year

Owner-occupied property

p10000	Filter information
p20000	Imputation flag for filter information
p0100x	Market value (imputation alternative $x = a, b, \dots, e$)
p02000	Imputation flag for market value
p0010x	Debts (imputation alternative $x = a, b, \dots, e$)
p00200	Imputation flag for debts
p00010	Individual share
p00020	Imputation flag for individual share
p0110x	Net market value ($p0100x - p0010x$; imputation alternative $x = a, b, \dots, e$)
p02200	Imputation flag for net market value
p0101x	Individual share of market value ($p0100x * p00010/100$; imputation alternative $x = a, b, \dots, e$)
p02020	Imputation flag for individual share of market value
p0011x	Individual share of debts ($p0010x * p00010/100$; imputation alternative $x = a, b, \dots, e$)
p00220	Imputation flag for individual share of debts
p0111x	Individual share of net market value ($(p0100x - p0010x) * p00010/100$; imputation alternative $x = a, b, \dots, e$)
p02220	Imputation flag for individual share of net market value

Other property

e10000	Filter information
e20000	Imputation flag for filter information
e0100x	Market value (imputation alternative $x = a, b, \dots, e$)
e02000	Imputation flag for market value
e00010	Individual share
e00020	Imputation flag for individual share
e0010x	Debts (imputation alternative $x = a, b, \dots, e$)
e00200	Imputation flag for debts
e0110x	Net market value ($e0100x - e0010x$; imputation alternative $x = a, b, \dots, e$)
e02200	Imputation flag for net market value
e0101x	Individual share of market value ($e0100x * e00010/100$; imputation alternative $x = a, b, \dots, e$)
e02020	Imputation flag for share of market value
e0011x	Individual share of debts ($e0010x * e00010/100$; imputation alternative $x = a, b, \dots, e$)
e00220	Imputation flag for individual share
e0111x	Individual share of net market value ($(e0100x - e0010x) * e00010/100$; imputation alternative $x = a, b, \dots, e$)
e02220	Imputation flag for individual share of net market value
e00001	Type: single-family house
e00002	Type: apartment building
e00003	Type: holiday home
e00004	Type: undeveloped real estate
e00005	Type: other property
e00006	Type: no answer
e00007	Number of properties

Financial Assets

f10000	Filter information
f20000	Imputation flag for filter information
f0100x	Market value (imputation alternative $x = a, b, \dots, e$)
f02000	Imputation flag for market value
f00010	Individual share
f00020	Imputation flag for individual share
f0101x	Individual share of market value ($f0100x * f00010 / 100$; imputation alternative $x = a, b, \dots, e$)
f02020	Imputation flag for individual share of market value

Private Insurances

i10000	Filter information
i20000	Imputation flag for filter information
i0100x	Market value (imputation alternative $x = a, b, \dots, e$)
i02000	Imputation flag for market value

Business Assets

b10000	Filter information
b20000	Imputation flag for filter information
b0100x	Market value (imputation alternative $x = a, b, \dots, e$)
b02000	Imputation flag for market value
b00001	Ownership status
b00002	Imputation flag for ownership status

Tangible Assets

t10000	Filter information
t20000	Imputation flag for filter information
t0100x	Market value (imputation alternative $x = a, b, \dots, e$)
t02000	Imputation flag for market value

Consumer Debts

c10000	Filter information
c20000	Imputation flag for filter information
c0100x	Market value (imputation alternative $x = a, b, \dots, e$)
c02000	Imputation flag for market value

Overall wealth

w0101x	Gross overall wealth ($p0101x + e0101x + f0101x + i0100x + b0100x + t0100x02$; imputation alternative $x = a, b, \dots, e$)
w02020	Imputation flag for gross overall wealth
w0011x	Overall debts ($p0011x + e0011x + c0100x$; imputation alternative $x = a, b, \dots, e$)
w00220	Imputation flag for overall debts
w0111x	Net overall wealth ($w0101x - w0011x$; imputation alternative $x = a, b, \dots, e$)
w02220	Imputation flag for net overall wealth

10.2 Variable list at the household level

As a matter of principle, the wealth-related variable names at the household level carry an “H” at the fifth digit identifying the unit of analysis (household). This fifth digit at the individual level carries information on the individual share which, due to aggregation at the household level, is not a relevant piece of information as such. Imputation flag variables are also aggregated across household members, i.e., household-level wealth information is considered to be imputed if for at least one household member imputed data from the respective component is available.

Identifiers

HHNRAKT Wave-specific household identifier
SVYYEAR Survey year

Property, primary residence

p100h0 HH filter information (max of p10000 over all HH-members)
p200h0 HH imputation flag for filter information
p010hx HH market value (sum of p0101x over all HH-members;
imputation alternative x = a,b,...,e)
p020h0 HH imputation flag for market value
p001hx HH debts (sum of p0011x over all HH-members;
imputation alternative x = a,b,...,e)
p002h0 HH imputation flag for debts
p011hx HH net value (p010Hx-p001Hx; imputation alternative x = a,b,...,e)
p022h0 HH imputation flag for net value

Other property

e100h0 HH filter information (max of e10000 over all HH-members)
e200h0 HH imputation flag for filter information
e010hx HH market value (sum of e0101x over all HH-members;
imputation alternative x = a,b,...,e)
e020h0 HH imputation flag for market value
e001hx HH debts (sum of e0011x over all HH-members;
imputation alternative x = a,b,...,e)
e002h0 HH imputation flag for debts
e011hx HH net value (e010Hx-e001Hx; imputation alternative x = a,b,...,e)
e022h0 HH imputation flag for net value

Financial assets

f100h0 HH filter information (max of f10000 over all HH-members)
f200h0 HH imputation flag for filter information
f010hx HH market value (sum of f0101x over all HH-members;
imputation alternative x = a,b,...,e)
f020h0 HH imputation flag for market value

Private insurances

i100h0	HH filter information (max of i10000 over all HH-members)
i200h0	HH imputation flag for filter information
i010hx	HH market value (sum of i0100x over all HH-members; imputation alternative x = a,b,...,e)
i020h0	HH imputation flag for market value

Business assets

b100h0	HH filter information (max of b10000 over all HH-members)
b200h0	HH imputation flag for filter information
b010hx	HH market value (sum of b0100x over all HH-members; imputation alternative x = a,b,...,e)
b020h0	HH imputation flag for market value

Tangible assets

t100h0	HH filter information (max of t10000 over all HH-members)
t200h0	HH imputation flag for filter information
t010hx	HH market value (sum of t0100x over all HH-members; imputation alternative x = a,b,...,e)
t020h0	HH imputation flag for market value

Consumer Debts

c100h0	HH filter information (max of c10000 over all HH-members)
c200h0	HH imputation flag for filter information
c010hx	HH market value (sum of c0100x over all HH-members)
c020h0	HH imputation flag for market value

Overall wealth

w010hx	HH gross overall wealth ($w010hx = p010Hx + e010Hx + f010Hx + i010Hx + b010Hx + t010Hx$; imputation alternative x = a,b,...,e)
w020h0	HH imputation flag for gross overall wealth
w001hx	HH overall debts: ($w001Hx = p001Hx + e001Hx + c010Hx$; imputation alternative x = a,b,...,e)
w002h0	HH imputation flag for overall debts
w011hx	HH net overall wealth ($w011Hx = w010Hx - w001Hx$; imputation alternative x = a,b,...,e)
w022h0	HH imputation flag for net overall wealth

10.3 How to perform analyses using multiply imputed values?

As described above, the setup of the fully imputed 2002 SOEP wealth data considers five variables for each wealth component per individual in a given household³³. While these five variables for a given wealth component carry identical information for individuals who fully

³³ Note that for children a value of 0 is assumed given that no wealth information is surveyed for children in the SOEP.

completed the respective wealth question, these five values will differ for observations with missing metric information due to the multiple imputation procedure.

Reshaping the “wide” data structure described above into a “long” structure (e.g., using the command *reshape* in the statistical software package Stata) and pooling the data from the five imputates will yield an easy to handle dataset. However, when using such multiply imputed data one has to consider that these five observations per person are not independent of each other and ignoring this issue will result in biased (underestimated) standard errors and possibly even point estimates.

A benchmark treatment of such data using the relevant results for the various (here five) imputates has been developed by Arthur B. Kennickell using multiply imputed data from the US Survey of Consumer Finances (SCF) (see Kennickell 1998). Stata code is available from <http://www.federalreserve.gov/pubs/oss/oss2/2004/codebk2004.txt> allowing computation of coefficients and standard errors, corrected for multiple imputation in various models (OLS, probit, etc.). For each independent variable of interest (including the constant term in a regression model), the output includes the corrected point estimate, standard error, t-statistic, and an indicator of the significance of the coefficient. The same strategy can be adopted for proper calculation of inequality indices or other moments of the income or wealth distribution.

References

- Ammermüller, Andreas, Andrea M. Weber und Peter Westerheide (2005): Die Entwicklung und Verteilung der Vermögens privater Haushalte unter besonderer Berücksichtigung des Produktivvermögens. Projektbericht an das Bundesministerium für Gesundheit und Soziale Sicherung. Mannheim: ZEW.*
- Bartzsch, Nikolaus and Stöss, Elmar (2006): Measuring German household debt: Financial accounts data and disaggregated survey data as complementary statistics. Prepared for the IFC conference in Basle, August 2006.*
- Becker, Irene; Frick, Joachim R.; Grabka, Markus M.; Hauser, Richard; Krause, Peter and Wagner, Gert G. (2002): A Comparison of the Main Household Income Surveys for Germany: EVS and SOEP. In: Hauser, Richard and Becker, Irene (Eds): Reporting on Income Distribution and Poverty. Perspectives from a German and a European Point of View. Heidelberg: Springer, S. 55-90.*
- Frick, Joachim R.; Grabka, Markus M. and Sierminska, Eva (2007): "Representative wealth data for Germany from the German SOEP: The impact of methodological decisions around imputation and the choice of the aggregation unit", DIW Discussion paper #672, Berlin: German Institute for Economic Research (DIW).*
- 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>*
- Schäfer, Andrea, and Jürgen Schupp (2006): Zur Erfassung der Vermögensbestände im Sozio-oekonomischen Panel (SOEP) im Jahr 2002, Data Documentation 11, Berlin: German Institute for Economic Research (DIW).*
- 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*
- Taylor, Marcia F. et al. (Eds.)(1998) British Household Panel Survey User Manual. Introduction, Technical Reports and Appendices, Colchester: University of Essex, ESRC.*

Appendix

Table 5: Comparison of the applied imputation implicates (individual level)³⁴

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Market value owner-occupied property (40.0%)	mean	50.891	50.674	50.674	50.833	50.781	50.771
	sd	93.489	92.718	93.239	93.565	93.489	93.300
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	152.701	154.000	153.702	155.000	152.931	153.667
	min	0	0	0	0	0	0
	max	3.750.000	3.750.000	3.750.000	3.750.000	3.750.000	3.750.000
	Gini	0,761	0,760	0,761	0,762	0,762	0,761
	HSCV	1,687	1,674	1,693	1,694	1,695	1,689
Debts related to owner-occupied property (22.0%)	mean	8.851	8.750	8.866	8.808	8.731	8.801
	sd	28.060	27.848	28.225	28.010	27.943	28.017
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	35.000	33.000	33.000	35.000	32.500	33.700
	min	0	0	0	0	0	0
	max	2.125.000	2.125.000	2.125.000	2.125.000	2.125.000	2.125.000
	Gini	0,899	0,899	0,899	0,899	0,899	0,899
	HSCV	5,025	5,064	5,067	5,057	5,122	5,067
Market value other property (12.2%)	mean	16270	16297	16321	16424	16624	16387
	sd	144987	142491	146012	143561	144701	144350
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	13.600.000	13.600.000	13.600.000	13.600.000	13.600.000	13.600.000
	Gini	0,961	0,961	0,961	0,960	0,961	0,961
	HSCV	39,701	38,217	40,013	38,196	37,880	38,801
Debts related to other property (6.1%)	mean	4.496	6426	4.527	5.726	4.488	5132
	sd	54.748	84061	55.094	68.777	54.890	63.514
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,981	0,984	0,981	0,983	0,981	0,981
	HSCV	74,133	85,541	74,055	72,124	74,779	76,126

³⁴ Weighted results based on 23,135 individual observations with a positive weighting factor (SPHRFAG); the percentage share of individuals holding the respective wealth component is given in brackets.

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Financial assets (44.9%)	mean	9.931	9.907	9.882	10.011	10.001	9.946
	sd	41.847	41.774	41.677	41.941	42.121	41.872
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	25.000	25.000	25.000	25.000	25.000	25.000
	min	0	0	0	0	0	0
	max	3.500.000	3.500.000	3.500.000	3.500.000	3.500.000	3.500.000
	Gini	0,833	0,833	0,833	0,834	0,833	0,833
	HSCV	8,878	8,888	8,892	8,775	8,868	8,860
Private pension (51.2%)	mean	9.037	9.038	9.198	9.152	9.116	9.108
	sd	36.727	36.523	37.757	37.341	37.283	37.126
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	23.369	23.970	24.030	23.557	23.442	23.674
	min	0	0	0	0	0	0
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,832	0,830	0,833	0,833	0,833	0,832
	HSCV	8,258	8,164	8,425	8,323	8,363	8,307
Business assets (5.2%)	mean	11.548	10.559	9.060	8.918	10.590	10.135
	sd	436.292	386.797	335.494	258.997	353.815	354.279
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	50.000.000	54.800.000	50.000.000	50.000.000	85.700.000	58.100.000
	Gini	0,994	0,993	0,992	0,992	0,993	0,993
	HSCV	713,698	670,923	685,576	421,746	558,051	609,999
Tangible assets (9.3%)	mean	1.400	1.404	1.370	1.419	1.469	1.412
	sd	18.096	17.140	16.980	16.914	18.459	17.518
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	3.000.000	3.000.000	3.000.000	3.000.000	3.000.000	3.000.000
	Gini	0,966	0,966	0,966	0,965	0,967	0,966
	HSCV	83,464	74,382	76,709	70,995	78,848	76,880
Consumer debts (11.1%)	mean	2.708	2.669	2.702	2.656	2.652	2.677
	sd	33.420	33.555	33.456	33.352	33.258	33.408
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	3.000	3.000	3.000	3.000	3.000	3.000
	min	0	0	0	0	0	0
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,965	0,965	0,965	0,965	0,965	0,965
	HSCV	76,141	79,000	76,642	78,843	78,625	77,850

Table 6: Comparison of the applied imputation implicates (individual level, for those owning the respective wealth component)³⁵

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Market value owner-occupied property (n=9,263)	mean	140.426	139.828	139.828	140.266	140.122	140.094
	sd	107.444	106.090	107.344	107.761	107.697	107.267
	p10	50.000	50.000	50.000	50.000	50.000	50.000
	p50	118.272	120.000	115.000	115.041	115.602	116.783
	p90	250.000	250.000	250.000	250.000	250.000	250.000
	min	4.000	4.000	4.000	4.000	4.000	4.000
	max	3.750.000	3.750.000	3.750.000	3.750.000	3.750.000	3.750.000
	Gini	0,340	0,338	0,341	0,342	0,342	0,341
	HSCV	0,293	0,288	0,295	0,295	0,295	0,293
Debts related to owner-occupied property (n=5,097)	mean	48.581	48.028	48.664	48.344	47.921	48.308
	sd	48.905	48.686	49.357	48.942	49.071	48.992
	p10	8.000	8.059	8.200	8.000	7.669	7.986
	p50	40.000	38.893	38.040	38.750	37.500	38.637
	p90	100.000	100.000	100.000	100.000	100.000	100.000
	min	240	240	240	240	240	240
	max	2.125.000	2.125.000	2.125.000	2.125.000	2.125.000	2.125.000
	Gini	0,444	0,443	0,447	0,443	0,446	0,445
	HSCV	0,507	0,514	0,514	0,512	0,524	0,514
Market value other property (n=2,811)	mean	162742	163015	163253	164286	166279	163.915
	sd	431835	423344	435104	426508	429654	429.289
	p10	15.000	15.000	14.754	14.368	15.000	14.824
	p50	87.500	82.242	87.500	88.993	89.488	87.144
	p90	300.000	312.500	300.000	320.000	306.775	307.855
	min	100	100	100	100	100	100
	max	13.600.000	13.600.000	13.600.000	13.600.000	13.600.000	13.600.000
	Gini	0,608	0,608	0,609	0,609	0,606	0,608
	HSCV	3,519	3,370	3,550	3,368	3,337	3,429
Debts related to other property (n=1,422)	mean	98.541	140.850	99.222	125.503	98.369	112497
	sd	237.621	368.815	239.098	297.822	238.405	276352
	p10	10.000	12.244	10.000	12.244	12.000	11298
	p50	52.118	60.000	56.000	60.000	52.500	56124
	p90	188.739	260.759	182.500	230.000	187.534	209906
	min	50	50	50	50	50	50
	max	5.000.000	6.892.764	5.000.000	5.000.000	5.000.000	5.378.552
	Gini	0,610	0,612	0,610	0,612	0,611	0,611
	HSCV	3,533	3,537	3,539	3,489	3,529	3,525

³⁵ Weighted results based on the population of individuals holding the respective wealth of debt component as well as a positive weighting factor (SPHRFAG); the number of those valid observations is given in brackets.

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Financial assets (n=10,381)	mean	23.088	23.035	22.976	23.275	23.251	23.125
	sd	61.379	61.274	61.133	61.489	61.778	61.412
	p10	3.000	3.000	3.000	3.000	3.000	3.000
	p50	10.000	10.000	10.000	10.000	10.000	10.000
	p90	50.000	50.000	50.000	50.000	50.000	50.000
	min	2500	2500	2500	2500	2500	2500
	max	3.500.000	3.500.000	3.500.000	3.500.000	3.500.000	3.500.000
	Gini	0,611	0,612	0,611	0,613	0,612	0,612
	HSCV	3,534	3,539	3,540	3,490	3,530	3,527
Private pension (n=11,852)	mean	19.036	19.038	19.375	19.278	19.203	19.186
	sd	51.490	51.182	52.971	52.365	52.293	52.060
	p10	1.500	1.500	1.500	1.500	1.500	1.500
	p50	8.080	8.428	8.567	8.264	8.234	8.315
	p90	44.706	42.770	44.933	43.000	40.903	43.262
	min	0	0	0	0	0	0
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,851	0,840	0,818	0,803	0,833	0,829
	HSCV	29,807	27,991	28,613	17,418	23,202	25,406
Business assets (n=1,194)	mean	272.125	248.826	213.504	210.145	249.568	238.833
	sd	2.101.969	1.862.556	1.615.809	1.240.840	1.700.790	1.704.393
	p10	3.221	3.000	3.000	3.507	2.500	3.045
	p50	43.643	40.931	43.845	47.784	50.000	45.241
	p90	400.000	400.000	394.517	400.000	400.000	398.903
	min	0	0	0	0	0	0
	max	50.000.000	54.800.000	50.000.000	50.000.000	85.700.000	58.100.000
	Gini	0,852	0,841	0,819	0,803	0,833	0,829
	HSCV	28,769	27,016	27,617	16,804	22,390	24,519
Tangible assets (n=2,148)	mean	16.595	16.651	16.243	16.819	17.417	16.745
	sd	60.245	56.820	56.356	55.966	61.328	58.143
	p10	3.000	3.000	3.000	3.000	3.000	3.000
	p50	8.000	8.394	8.000	8.297	8.000	8.138
	p90	30.000	30.000	30.000	31.403	32.808	30.842
	min	2500	2500	2500	2500	2500	2500
	max	3.000.000	3.000.000	3.000.000	3.000.000	3.000.000	3.000.000
	Gini	0,599	0,591	0,595	0,588	0,607	0,596
	HSCV	6,585	5,819	6,015	5,533	6,199	6,030
Debts from consumer credits (n=2,565)	mean	25.428	25.065	25.373	24.938	24.902	25.141
	sd	99.564	100.074	99.692	99.461	99.174	99.593
	p10	3.500	3.500	3.500	3.500	3.500	3.500
	p50	9.255	9.000	9.000	9.000	9.000	9.051
	p90	46.000	44.000	46.513	42.500	49.000	45.603
	min	2500	2500	2500	2500	2500	2500
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,675	0,674	0,675	0,673	0,671	0,674
	HSCV	7,662	7,966	7,715	7,949	7,926	7,844

Table 7: Comparison of the applied imputation implicates (household level)³⁶

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Market value owner-occupied property (47.4%)	mean	88.632	88.273	88.323	88.570	88.571	88.474
	sd	147.212	146.603	146.715	146.993	147.371	146.978
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	267.157	260.000	262.779	270.000	266.159	265.219
	min	0	0	0	0	0	0
	max	7.500.000	7.500.000	7.500.000	7.500.000	7.500.000	7.500.000
	Gini	0,725	0,725	0,725	0,725	0,726	0,725
	HSCV	1,379	1,379	1,379	1,377	1,384	1,380
Debts related to owner-occupied property (25.0%)	mean	15.590	15.558	15.634	15.548	15.472	15.560
	sd	49.664	49.616	49.910	49.607	49.682	49.696
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	60.641	60.000	60.000	60.000	60.000	60.128
	min	0	0	0	0	0	0
	max	4.250.000	4.250.000	4.250.000	4.250.000	4.250.000	4.250.000
	Gini	0,891	0,891	0,891	0,891	0,892	0,891
	HSCV	5,074	5,084	5,095	5,089	5,155	5,099
Market value other property (15.9%)	mean	29.819	30.068	30.207	30.672	30.585	30.270
	sd	240.060	237.026	241.396	239.666	240.262	239.682
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	39.000	38.750	38.537	39.327	40.000	39.122
	min	0	0	0	0	0	0
	max	18.000.000	18.000.000	18.000.000	18.000.000	18.000.000	18.000.000
	Gini	0,953	0,953	0,953	0,953	0,953	0,953
	HSCV	32,401	31,066	31,928	30,524	30,852	31,354
Debts related to other property (8.0%)	mean	8.606	13.837	8.690	12.874	8.755	10.552
	sd	92.082	138.548	92.776	129.315	92.646	109.073
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000
	Gini	0,978	0,981	0,977	0,981	0,978	0,979
	HSCV	57,236	50,117	56,974	50,437	55,984	54,150

³⁶ Weighted results based on 12,308 households with a positive weighting factor (SHHRFAG); the percentage share of all households possessing the respective component is given in brackets.

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Financial assets (53.3 %)	mean	17.449	17.276	17.326	17.411	17.427	17.377
	sd	68.172	66.906	68.216	66.607	67.393	67.459
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	42.000	42.562	41.468	40.903	42.500	41.887
	min	0	0	0	0	0	0
	max	6.000.000	6.000.000	6.000.000	6.000.000	6.000.000	6.000.000
	Gini	0,808	0,808	0,808	0,807	0,808	0,808
	HSCV	7,630	7,497	7,749	7,316	7,476	7,534
Private pension (61.8%)	mean	15.437	15.444	15.706	15.615	15.589	15.558
	sd	60.531	60.393	61.946	61.419	61.316	61.121
	p10	0	0	0	0	0	0
	p50	2.000	2.000	2.000	2.000	1.860	1.972
	p90	40.000	40.000	40.000	40.000	40.000	40.000
	min	0	0	0	0	0	0
	max	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000
	Gini	0,804	0,803	0,805	0,805	0,806	0,805
	HSCV	7,687	7,645	7,777	7,734	7,735	7,716
Business assets (8.6%)	mean	21.326	19.708	17.284	16.919	20.186	19.085
	sd	661.999	580.117	518.759	439.329	592.289	558.499
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	0	0	0	0	0	0
	min	0	0	0	0	0	0
	max	50.800.000	56.500.000	50.000.000	50.000.000	85.700.000	58.600.000
	Gini	0,991	0,991	0,989	0,989	0,990	0,990
	HSCV	481,742	433,212	450,357	337,122	430,424	426,571
Tangible assets (13.0%)	mean	2.480	2.495	2.447	2.508	2.590	2.504
	sd	28.109	28.077	27.374	27.446	27.917	27.784
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	3.000	3.000	3.000	3.000	3.000	3.000
	min	0	0	0	0	0	0
	max	3.200.795	3.312.095	3.257.147	3.386.519	3.054.915	3.242.294
	Gini	0,959	0,959	0,960	0,959	0,960	0,959
	HSCV	64,180	63,212	62,529	59,855	58,069	61,569
Consumer debts (16.1%)	mean	4.666	4.609	4.694	4.604	4.601	4.634
	sd	45.338	45.595	45.646	45.244	44.967	45.358
	p10	0	0	0	0	0	0
	p50	0	0	0	0	0	0
	p90	7.407	7.000	7.000	7.000	7.419	7.165
	min	0	0	0	0	0	0
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,954	0,955	0,955	0,955	0,954	0,955
	HSCV	47,190	48,927	47,264	48,273	47,744	47,880

Table 8: Comparison of the applied imputation implicates (household level; only those owning the respective wealth component)³⁷

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Market value owner-occupied property (n=5,830)	mean	218.930	218.043	218.166	218.778	218.779	218.539
	sd	158.127	157.457	157.613	157.749	158.617	157.913
	p10	85.710	85.819	87.500	85.000	84.806	85.767
	p50	200.000	198.656	194.291	199.859	198.254	198.212
	p90	390.000	390.068	385.000	387.101	396.000	389.634
	min	4.000	4.000	4.000	4.000	4.000	4.000
	max	7.500.000	7.500.000	7.500.000	7.500.000	7.500.000	7.500.000
	Gini	0,320	0,320	0,319	0,320	0,322	0,321
	HSCV	0,261	0,261	0,261	0,259	0,263	0,261
Debts related to owner-occupied property (n=3,079)	mean	80.705	80.540	80.934	80.490	80.094	80.553
	sd	86.690	86.670	87.246	86.680	87.199	86.897
	p10	14.000	15.000	15.000	13.502	14.000	14.300
	p50	64.564	62.500	62.500	64.000	62.345	63.182
	p90	168.894	167.500	167.500	168.456	166.034	167.677
	min	480	480	480	480	371	458
	max	4.250.000	4.250.000	4.250.000	4.250.000	4.250.000	4.250.000
	Gini	0,436	0,436	0,436	0,435	0,438	0,436
	HSCV	0,577	0,579	0,581	0,580	0,592	0,582
Market value other property (n=1,956)	mean	241.138	243.152	244.267	248.033	247.326	244.783
	sd	644.386	634.561	647.383	640.896	642.936	642.032
	p10	17.694	16.748	17.500	18.023	20.000	17.993
	p50	125.000	125.000	127.500	126.000	125.000	125.700
	p90	454.650	485.944	469.973	492.779	471.984	475.066
	min	100	100	100	100	100	100
	max	18.000.000	18.000.000	18.000.000	18.000.000	18.000.000	18.000.000
	Gini	0,619	0,621	0,618	0,622	0,615	0,619
	HSCV	3,568	3,403	3,510	3,336	3,377	3,439
Debts related to other property (n=983)	mean	153.619	247.008	155.132	229.815	156.277	188.370
	sd	359.437	534.142	362.002	498.866	360.939	423.077
	p10	15.000	20.000	16.500	20.000	17.500	17.800
	p50	80.000	90.000	80.000	90.000	80.000	84.000
	p90	300.000	600.000	300.000	481.922	300.000	396.384
	min	50	50	50	50	50	50
	max	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000
	Gini	0,602	0,637	0,600	0,671	0,599	0,622
	HSCV	2,734	2,335	2,719	2,353	2,664	2,561

³⁷ Weighted results based on the population of household holding the respective wealth of debt component as well as a positive weighting factor (SHHRFAG); the number of those valid observations is given in brackets.

		Implicate 1	Implicate 2	Implicate 3	Implicate 4	Implicate 5	Overall Mean
Financial assets (n=6,560)	mean	35.314	34.968	35.068	35.240	35.273	35.172
	sd	93.678	91.881	93.791	91.387	92.540	92.655
	p10	4.626	4.898	4.866	4.715	4.641	4.749
	p50	16.683	16.000	16.890	16.332	16.948	16.571
	p90	75.000	71.243	73.897	73.722	75.000	73.772
	min	380	380	380	380	380	380
	max	6.000.000	6.000.000	6.000.000	6.000.000	6.000.000	6.000.000
	Gini	0,612	0,611	0,610	0,610	0,611	0,611
	HSCV	3,517	3,451	3,575	3,361	3,440	3,469
Private pension (n=7,609)	mean	28.080	28.092	28.570	28.404	28.356	28.300
	sd	79.437	79.242	81.320	80.616	80.481	80.219
	p10	2.000	2.000	2.000	2.000	2.000	2.000
	p50	12.000	12.000	12.000	12.000	12.000	12.000
	p90	61.934	61.400	62.181	61.400	60.330	61.449
	min	10	10	10	10	10	10
	max	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000	8.000.000
	Gini	0,643	0,641	0,646	0,646	0,647	0,645
	HSCV	4,001	3,978	4,050	4,027	4,027	4,017
Business assets (n=1,059)	mean	330.051	304.997	267.496	261.834	312.404	295.356
	sd	2.585.766	2.263.997	2.025.193	1.710.393	2.311.374	2.179.345
	p10	5.000	4.800	5.000	5.000	4.800	4.920
	p50	50.000	50.000	50.000	50.000	50.000	50.000
	p90	400.000	413.897	403.212	425.136	500.000	428.449
	min	1	1	1	1	1	1
	max	50.800.000	56.500.000	50.000.000	50.000.000	85.700.000	58.600.000
	Gini	0,863	0,854	0,837	0,826	0,851	0,846
	HSCV	30,660	27,525	28,632	21,316	27,344	27,095
Tangible assets (n=1,597)	mean	22.231	22.376	21.934	22.408	23.213	22.432
	sd	81.515	81.381	79.314	79.404	80.677	80.458
	p10	3.000	3.000	3.000	3.000	3.000	3.000
	p50	10.000	10.000	10.000	10.000	10.000	10.000
	p90	41.008	43.125	41.935	45.000	46.179	43.449
	min	2.500	2.500	2.500	2.500	2.500	2.500
	max	3.200.795	3.312.095	3.257.147	3.386.519	3.054.915	3.242.294
	Gini	0,635	0,631	0,636	0,628	0,640	0,634
	HSCV	6,717	6,609	6,533	6,235	6,035	6,426
Consumer debts (n=1,976)	mean	32.466	32.065	32.661	32.033	32.012	32.247
	sd	115.775	116.571	116.565	115.622	114.872	115.881
	p10	4.000	4.000	4.000	4.000	4.000	4.000
	p50	11.679	11.583	12.000	12.000	12.000	11.852
	p90	60.000	55.000	55.000	55.000	59.284	56.857
	min	2.500	2.500	2.500	2.500	2.500	2.500
	max	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000	5.000.000
	Gini	0,683	0,685	0,684	0,684	0,681	0,684
	HSCV	6,354	6,604	6,365	6,510	6,434	6,453

Table 9: Comparison of observed values and values edited/imputed³⁸ by Frick/Grabka/Marcus (2007) and Schupp/Schaefer (2006) (unweighted results)

	Frick/Grabka/Marcus (FGM2007)			Schupp/Schaefer (SS2006)		
	N	mean (euros)	median (euros)	N	mean (euros)	median (euros)
Market value owner-occupied property						
observed	6.596	152.360	125.000	8.004	153.595	125.000
imputed/edited	3.293	124.866	104.672	1.104	142.585	151.114
total	9.889	143.204	116.598	9.108	152.261	125.000
Market value other property						
observed	2.229	189.501	100.000	2.361	206.530	100.000
imputed/edited	799	197.985	76.886	453	169.200	213.946
total	3.028	191.740	90.856	2.814	200.374	110.000
Financial assets						
observed	7.691	28.066	10.000	8.365	27.769	10.000
imputed/edited	3.449	19.732	10.354	1.822	25.001	26.850
total	11.140	25.486	10.000	10.187	27.274	15.000
Private pension						
observed	8.309	24.775	10.000	8.310	24.772	10.000
imputed/edited	4.592	17.011	7.317	3.308	20.852	21.786
total	12.901	22.011	8.747	11.618	23.656	13.000
Business assets						
observed	833	284.290	50.000	833	284.290	50.000
imputed/edited	428	471.856	55.322	350	226.955	304.757
total	1.261	347.953	50.000	1.183	267.327	100.000
Tangible assets						
observed	1.618	22.088	8.000	1.618	22.088	8.000
imputed/edited	688	18.843	10.261	592	21.352	19.219
total	2.306	21.120	9.514	2.210	21.891	10.000
Debts related to owner-occupied property						
observed	3.525	53.108	42.500	4.274	54.906	43.000
imputed/edited	1.968	46.566	34.698	62	55.855	55.855
total	5.493	50.764	39.963	4.336	54.919	45.000
Debts related to other property						
observed	1.152	113.780	56.000	1.216	129.773	43.000
imputed/edited	397	214.326	81.607	6	115.5601	115.5601
total	1.549	139.549	61.990	1.222	129.703	57.625
Consumer debts						
observed	2.244	26.545	9.650	2.244	26.545	9.650
imputed/edited	502	26.620	11.828	366	25.921	24.376
total	2.746	26.559	10.000	2.610	26.457	10.000

³⁸ Editing/imputation applies to either the market value or the individual share.

Table 10: Influence of the imputation/editing process (individual level, weighted)

Wealth component	Value <i>before</i> imputation/editing ³⁹			Value <i>after</i> imputation/editing ⁴⁰			Share of imputed wealth (in %) ⁴¹
	Mean (euros)	Volume in Mio.	Share in %	Mean (euros)	Volume in Mio.	Share in %	
Property (owner-occupied)	47.314	2.829.818	56,4	50.771	3.421.346	51,9	17,3
Property (other)	13.695	850.818	17,0	16.388	1.104.348	16,8	23,0
Financial assets	9.052	501.477	10,0	9.947	670.340	10,2	25,2
Private pension	7.870	425.161	8,5	9.108	613.777	9,3	30,7
Commercial enterprise	5.513	344.622	6,9	10.135	682.977	10,4	49,5
Tangible assets	1.046	64.067	1,3	1.413	95.231	1,4	32,7
Gross overall wealth	84.490	5.015.963	100,0	97.762	6.588.019	100,0	23,9
Debts owner-occupied property	8.035	484.269	55,4	8.801	593.092	53,0	18,3
Debts other property	3.855	241.597	27,6	5.133	345.900	30,9	30,2
Debts from consumer credits	2.381	148.468	17,0	2.678	180.435	16,1	17,7
Overall charges	14.271	874.334	100,0	16.612	1.119.427	100,0	21,9
Net overall wealth	70.219	4.141.629	--	81.150	5.468.592	--	24,3

³⁹ Based on observations with valid information on all relevant inputs (metric values and individual share) including those with original filter information equal 'no' (i.e. value of component is 0).

⁴⁰ Values refer to the mean of the five versions.

⁴¹ Calculated as [(Volume after imputation – volume before imputation)/volume after imputation]

Table 11: Influence of the imputation/editing process (household level, weighted)

Wealth component	Value <i>before</i> imputation/editing ⁴²			Value <i>after</i> imputation/editing ⁴³			Share of imputed wealth (in %) ⁴⁴
	Mean (euros)	Volume in Mio.	Share in %	Mean (euros) ⁴⁵	Volume in Mio.	Share in %	
Property (owner-occupied)	70.541	2.758.532	54,9	88.474	3.459.783	51,1	20,3
Property (other)	22.354	874.147	17,4	30.271	1.183.744	17,5	26,2
Financial assets	12.658	495.009	9,9	17.378	679.585	10,0	27,2
Private pension	10.683	417.754	8,3	15.558	608.408	9,0	31,3
Commercial enterprise	10.480	409.831	8,2	19.085	746.308	11,0	45,1
Tangible assets	1.687	65.963	1,3	2.505	97.953	1,4	32,7
Gross overall wealth	128.403	5.021.236	100,0	173.058	6.775.781	100,0	25,9
Debts owner-occupied property	12.020	470.063	54,4	15.560	608.490	50,6	22,7
Debts other property	6.383	249.594	28,9	10.553	412.677	34,3	39,5
Debts from consumer credits	3.702	144.758	16,7	4.635	181.263	15,1	20,1
Overall charges	22.105	864.415	100,0	28.785	1.202.430	100,0	23,2
Net overall wealth	106.298	4.156.821	--	144.273	5.573.351	--	25,4

⁴² Based on observations with valid information on all relevant inputs (metric values and individual share) including those with original filter information equal 'no' (i.e. value of component is 0)..

⁴³ Values refer to the mean of the five versions.

⁴⁴ Calculated as [(Volume after imputation – volume before imputation) / volume after imputation]

⁴⁵ The mean overall value of the household's wealth includes eventual imputation of PUNR.

Table 12: The effect of editing and imputation on wealth inequality (individual level, weighted)

	Total population			Population with component		
	Observed ¹	Final ²	% change due to imputation ³	Observed ¹	Final ²	% change due to imputation ³
Owner-occupied property (PR)						
Gini	0,794	0,761	-4,2	0,353	0,341	-3,4
HSCV	2,021	1,688	-16,5	0,302	0,293	-3,0
Other property (IR)						
Gini	0,967	0,961	-0,6	0,599	0,608	1,5
HSCV	55,778	38,802	-30,0	4,073	3,429	-15,8
Financial assets						
Gini	0,870	0,833	-4,3	0,637	0,612	-3,9
HSCV	12,141	8,859	-27,0	4,011	3,526	-12,1
Private pensions						
Gini	0,869	0,832	-4,3	0,656	0,644	-1,8
HSCV	12,083	8,307	-31,3	4,302	3,657	-15,0
Business assets						
Gini	0,994	0,993	-0,1	0,783	0,823	5,1
HSCV	938,845	609,999	-35,0	24,884	24,519	-1,5
Tangible assets						
Gini	0,977	0,966	-1,1	0,626	0,596	-4,8
HSCV	133,013	76,880	-42,2	7,812	6,030	-22,8
Owner-occupied property debts						
Gini	0,916	0,899	-1,9	0,453	0,445	-1,8
HSCV	6,991	5,067	-27,5	0,648	0,514	-20,7
Other property debts						
Gini	0,984	0,982	-0,2	0,592	0,609	2,9
HSCV	108,558	76,125	-29,9	3,699	2,996	-19,0
Other debts						
Gini	0,969	0,965	-0,4	0,683	0,674	-1,3
HSCV	100,788	77,851	-22,8	9,250	7,844	-15,2

Note:

¹ Only observations with valid data on personal share and metric value are included.² After editing and imputation³ (final observations)/observations

Table 13: Set-up of and covariates used in regression model

		Owner-occupied property					Other property					Financial assets			Private pensions		Business assets			Tangible assets		Consumer debt			
		Filter	Individual Share	Debt (Y/N)	Market Value	Debt	Filter	Individual Share	Debt (Y/N)	Market Value	Debts	Filter	Individual Share	Market Value	Filter	Market Value	Filter	Individual Share	Market Value	Filter	Market Value	Filter	Market Value	Filter	Consumer credit
Legend:																									
X		Independent variable used in respective regression model.																							
1		Variable is used only as selection variable in the Heckman selection model																							
2		The logarithm of the variable is used																							
3		For owner-occupied property, the original value (prior to the regression based imputation) is used (see Section 3.3)																							
(X)		Variable is excluded from the logistic regression model for PUNR																							
		For all dependent metric variables the respective logarithms are used.																							
Variable	Description (original variable name)	this information has been completely constructed by the editing process (see Sections 3.1 and 3.2)																							
How dwelling was acquired	3 dummies: acquired by purchase, inheritance or newly built? (SERWERB)																								
Age	Age of the respondent. Missing values were estimated on the basis of the age of other household members and the relationship to the head of household (GEBJAHR)																								
Age of house	7 dummies: Year of construction: before 1918, 1918-1948, 1949-1971, 1972-1980, 1981-1990, 1991-2000, 2001 and later (SBAUJ)?																								
Age ²	Age of individual squared (GEBJAHR)																								
Savings account	Dummy: Household holds savings account (yes=1) (SH4301)																								
Building Contract	Dummy: Household holds a contract with a savings and loans society (yes=1) (SH4302)																								
Capia02	4 dummies for the categorical CAPI-information on the market value of owner-occupied property (<150.000, <200.000, <400.000, >400.000) (SP85AC1*)																								
Capia03	4 dummies for the categorical CAPI-information on the level of debts of owner-occupied property (<20.000, <50.000, <150.000, >150.000) (SP85AC2*)																								
Capib08	4 dummies for the categorical CAPI-information on the market value of other property (<20.000, <150.000, <500.000, >500.000) (SP85BC3*)																								
Capib11	4 dummies for the categorical CAPI-information on the debt value for other property (<10.000, <100.000, <250.000, >250.000) (SP85BC4*)																								
Capic02	4 dummies for the categorical CAPI-information on the value of financial assets (<5.000, <20.000, <100.000, >100.000) (SP85CC5*)																								
Capid02	4 dummies for the categorical CAPI-information on the value of private pensions (<5.000, <20.000, <100.000, >100.000) (SP85DC6*)																								

Capie03	4 dummies for the categorical CAPI-information on the value of business assets (<5.000, <50.000, <500.000, >500.000) (SP85EC7*)																X			
Capif02	4 dummies for the categorical CAPI-information on the value of tangible assets (<5.000, <20.000, <100.000, >100.000) (SP85FC8*)																			X
Capig02	4 dummies for the categorical CAPI-information on the value of consumer credits (<5.000, <10.000, <50.000, >50.000) (SP85GC9*)																			X ²
Children	Dummy: children younger than sixteen in the household (yes=1) (SH58)				X					X	X		X		X					X
Civil servant	Dummy: civil servant (yes=1) (SP4005)	X ¹	X ¹				X ¹	X ¹			X ¹	(X)	X	(X)		X ¹		X ¹		X ¹
Condition of house	2 dummies. Dwelling is in a good condition (yes=1); Dwelling needs major refurbishment (yes=1) (SRENOV)	X	X																	
Credit	Dummy: household raised a consumer credit (yes=1) (SH5001)				X			X			X					X			X	X
Credit value	Monthly amount of loan repayment (SH5002)																			X
Debts other property	Debts related to other property (SP85B11/SP85B12 after editing and imputation)					X ²														
Debts owner-occupied property	Debts related to owner-occupied property (SP85A03 edited/imputed; the first of the imputed versions is taken) ³	X ²			X ²											X ²				
Dishwasher	Dummy: Dishwasher in the household (yes=1) (SH5217)	X																		
District type	10 dummies on categorical information of the district's size (SBIK)	X																		
Dividend	Dividend income in the household, metric information from SH4401 are logarithmized, categorical information from SH4402 are recoded into 6 dummies, one for each category (<250, <1.000, <2.500, <5.000, <10.000, >10.000)									X		X								
Double	Dummy: household owns both, life assurance and building contract (yes=1)											X								
Dwelling satisfaction	Satisfaction with the dwelling. For those without a valid info. the mean of all other household members was used or (if no household member gave a valid information to this satisfaction question) a random number between 0 and 10 was taken	X																		
Education	Years of education. Those who are still in education are assigned the minimum of seven years. (SBILZEIT)	X		(X)	X		X	X	X	(X)						(X)	X			
Enterprise	Dummy: household owns a commercial enterprise (yes=1) (SH4306)	X										X		X						X
Equipment	2 dummies: household with garden / balcony (SAUS5/SAUS7)	X																		
Estimated rent	Estimation of monthly rent by owners if they had to rent their dwelling (SH3802)	X ²																		
Financial worries	Dummy: At least some concerns about finances (yes=1) (SP11302)									X										
Full-time	Dummy: full-time-employed (yes=1) (SP15)			(X)				(X)		(X)	(X)	(X)							(X)	
German	Dummy: born in Germany (yes=1) (GERMBORN)						(X)			(X)	(X)	(X)				(X)		(X)		

Household income	Annual post-government household income in euros (i1110202)		X ²		X ²		X ²	X ²	X ²		X ²	X ²	X ²		X ²	X ²	X ²	X ²	X ²
Inheritance of property	Dummy: Did the person ever inherit property prior to 2001 (yes=1) (RP108A03)			X	X		X	X											
Inheritance	Dummy: Did the person ever inherit prior to 2001 (yes=1) (RP108)			X	X								X						
Inheritance	Dummy: Household received inheritance/other windfall profits in the previous year (yes=1) (SH4501)			(X)				X		X						X			
Job tenure	Job tenure of respondent in years (SERWZEIT)											X		X					(X)
Labor earning	Annual individual labour earnings in euros (i1111002)			(X ²)									(X ²)						(X ²)
Life assurance	Dummy: Life assurance in household (yes=1) (SH4303)									X	X								
Dissatisfaction with life	Dummy: Individual is unhappy with his/her life (SP13501<6)	X ¹	X ¹				X ¹	X ¹			X ¹	X ¹	X ¹			X ¹		X ¹	X ¹
Satisfaction with life	Dummy: Individual is happy with his/her life (SP13501>=9)	X ¹	X ¹				X ¹	X ¹			X ¹	X ¹	X ¹			X ¹		X ¹	X ¹
Standard of living	Satisfaction with standard of living (SP0110)											X					(X)		
Mainten. owner-occupied property	Maintenance costs for the owner-occupied property in euros (SH33)																		(X)
Maintenance cost	Yearly maintenance costs for other properties in euros (SH4201)						X ²	X ²											
Marital status	5 dummies for married, married but separated, single, divorced, widowed (SFAMSTD)				X				(X)	X									(X)
Missing	Dummies for all those variables where missing values exist: missing or valid information	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Monthly savings	Dummy: Household has monthly savings (yes=1) (SH5101)				X		X	X			X		X			X		X	X
New car	Dummy: Purchase of a new car in the last 12 months (yes=1) (SH5202)																		(X) X
No debt owner-occupied property	Dummy: Debts for owner-occupied property (no debt=1) (SP85A03, edited/imputed)	X		X			X			X						X			
No job degree	Dummy: no vocational degree (yes=1) (SPBBIL03)															(X)			
No partner	Dummy: no partner within the household (yes=1) (PARTNR02)			X	X				X	X		X	X			X	X	X	X
No paym. to others	Dummy: no payments/support to persons outside the household (yes=1) (SP13421)									X									
Number of other property items	3 dummies for the number of other property: (0 items/2-3 items/4+ items) (SP85B07)						X	X											
Occupancy	Year moved into dwelling (SEINZUG)		X				X ²	X ²											X ²
Owner	Dummy: Does the person have own property (yes=1) (SP85A01, edited/imputed)			X	X		X	X		X	X	X			X		X		X

Partner's filter	Dummy: Does the partner possess the respective wealth component (yes=1) (SP85xx01)			X	X				X	X	X	X	X	X	X	X	X	X	X
Partner's share	Partner's share of the respective wealth component (metric for other properties, 2 dummies (for 100% and 50%) for financial assets) (SP85B09/B10 resp. SP85C03/SP85C04)				X					X									
Partner's value	Partner's value of the respective wealth component (SP85C02, SP85D02, SP85F02, SP85G02)										X ²		X ²					X ²	X ²
Part-time	Dummy: part-time-employed (yes=1) (SP15)			(X)									(X)					(X)	
Paym. dwelling(m)	Monthly loan payments for owner-occupied property in euros (SH32)		X ²				X ²	X ²			X ²						X		
Paym. dwelling	Dummy: Monthly payments for the owner-occupied property yes=(1) (SH31)			X									X						
Payments other property	Annual loan payments for other property in euros (SH4202)						X ²												
Private health insurance	Dummy: Does the individual have a private health insurance (yes=1) (SP103)			(X)			X					X	(X)			(X)		(X)	
Old-age provisions	5 dummies: Interest in building-up private old-age provision (very strong/strong/medium/less/not at all) (SP81)							(X)		X	(X)	X				(X)			
Public sector	Dummy: Individual works in the public sector (yes=1) (OEFFD02)												X						X
Region	97 dummies: <i>Raumordnungsregion</i> (SROR)	X	X																
Rent income	Dummy: Household receives income from renting & leasing (yes =1) (SH40)	X		X	X		X					X	X		X				
Rent income 2003	Dummy: Household receives income from renting & leasing in the following year (yes=1) (TH38)			X															
Rent income (met)	Household income from renting & leasing in the previous year in euros (SH41)				X ²		X ²								X ²				X ²
Rent level	6 dummies identifying regional level of rent (<i>Mietstufe</i>)	X																	
Residential area	3 dummies on type of residential area: "predominantly old houses / predominantly new houses / other" (SWUM3)	X																	
Life Satisfaction	Satisfaction with life in general (SP13501)																		(X)
Income Satisfaction	11 dummies for the categories of satisfaction with household income (SP0104)							(X)											
Self-employed	Dummy: individual is self-employed (yes=1) (SP4002)																	(X)	X
Self-employment	6 Dummies for self-employment status: farmer, free-lancer, without co-workers, with <10 co-workers, with >9 co-workers, helper in family business (SP4002)												X	X	X				
Sex	Dummy: female = 1 (SEX)			X	X	X		X		X	X	X	X			X	X	X	
Share owner-occupied property	2 dummies: share of owner-occupied property after editing/imputation is 100% and 50%, respectively (SP85A06, edited/imputed)				X				X										

Size housing unit	Size of the housing unit in square meters. For missing values, the mean of those with the same number of rooms resp. the same number of household members (if the information on the amount of rooms was also missing) was imputed (SWOHNFL)		X ²										X ²					X ²		X ²
Size of household	3 dummies for size of household (one person / two or three / 4+ persons) (SHHGR)				X			X	X	X	X	X		X		X	X	X	X	X
Social assistance	Dummy: Household received social assistance in the previous year (yes=1) (SH4610)							X												
Sole owner of enterprise	Dummy: Sole owner of the enterprise (yes=1) (SP85E02 + imputation)													X						
State contributions	4 dummies: Importance of public contributions for private provision (very important / important / less important / not all important) (SP84)							(X)												
Fixed interest securities	Dummy: Household owns stocks (yes=1) (SH4304, SH4305)							X		X										
Type of house	8 dummies: Type of house (farm house, one- or two-family house, one- or two-family row house, 3-4 unit building, 5-8 unit building, 9- or more unit building, other) (SWUM1)	X	X																	
Type of property	5 dummies: house/apartment, multiple family/apartment house, holiday home, undeveloped land, other property (SP85B02-B06)			X		X	X													
Unemployed	Dummy: Individual is unemployed (yes=1) (SP10)										X				X					
Value other property	Market value of other property in euros (SP85B08 after editing and imputation)						X ²													
Value owner-occupied property	Market value of owner-occupied property (SP85A02 edited/imputed; the first of the imputed versions is taken) ³		X ²		X	X ²							X ²	X ²						X ²
West	Dummy: West Germany (yes=1) (SBULA)				X	X		X	X	X	X	X		X		X		X		X
R ² /Pseudo-R ² * 100	(Values in brackets relate to the logistic regression model for PUNR)	52	59	46 (45)	44	17	49	50	42 (40)	66	63	38 (34)	50 (34)	62	15	42	30 (29)	51 (36)	38	54