

SOEPpapers
on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel study at DIW Berlin

790-2015

**Longitudinal Wealth Data
and Multiple Imputation**

An Evaluation Study

Christian Westermeier and Markus M. Grabka

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:

Jan **Goebel** (Spatial Economics)

Martin **Kroh** (Political Science, Survey Methodology)

Carsten **Schröder** (Public Economics)

Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics)

Denis **Gerstorff** (Psychology, DIW Research Director)

Elke **Holst** (Gender Studies, DIW Research Director)

Frauke **Kreuter** (Survey Methodology, DIW Research Fellow)

Frieder R. **Lang** (Psychology, DIW Research Fellow)

Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)

Thomas **Siedler** (Empirical Economics)

C. Katharina **Spieß** (Education and Family Economics)

Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)

DIW Berlin

Mohrenstrasse 58

10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de



Longitudinal Wealth Data and Multiple Imputation

An Evaluation Study

Christian Westermeier*, Markus M. Grabka, DIW Berlin[†]

Abstract

Statistical Analysis in surveys is generally facing missing data. In longitudinal studies for some missing values there might be past or future data points available. The question arises how to successfully transform this advantage into improved imputation strategies. In a simulation study the authors compare six combinations of cross-sectional and longitudinal imputation strategies for German wealth panel data. The authors create simulation data sets by blanking out observed data points: they induce item non response by a missing at random (MAR) and two differential nonresponse (DNR) mechanisms. We test the performance of multiple imputation using chained equations (MICE), an imputation procedure for panel data known as the row-and-column method and a regression prediction with correction for sample selection. The regression and MICE approaches serve as fallback methods, when only cross-sectional data is available. The row-and-column method performs surprisingly well considering the cross-sectional evaluation criteria. For trend estimates and the measurement of inequality, combining MICE with the row-and-column technique regularly improves the results based on a catalogue of six evaluation criteria including three separate inequality indices. As for wealth mobility, two additional criteria show that a model based approach such as MICE might be the preferable choice. Overall the results show that if the variables, which ought to be imputed, are highly skewed; the row-and-column technique should not be dismissed beforehand.

Key words:

Panel data, SOEP survey, evaluation, simulation, missing at random, item non-response

* Corresponding author, contact at cwestermeier@diw.de, +49 30 89789-223.

[†] The authors gratefully acknowledge funding from the *Hans Böckler Foundation*.

1 Introduction

Large-scale surveys are usually facing missing data, which poses problems for researchers and research infrastructure providers alike. In longitudinal studies for some missing values there might be past or future data points available. The question arises how to successfully transform this advantage into improved imputation strategies. Single imputation proves to have undesired properties, because the uncertainty reflected by the respective parameters based on one single stochastic imputation is likely to be biased downwards, since the estimators treat the imputed values as if they were actually observed ones (Rubin, 1987, 1996).¹ Multiple imputation addresses this issue. Our study examines the performance of several multiple imputation methods for the adjustment for item-non response (INR) in wealth panel data. Wealth is considered a sensitive information that is usually collected with rather high nonresponse rates compared to less sensitive questions such as pure demographic variables like age, sex, migration status (e.g. Riphahn & Serfling, 2005, Frick, Grabka, & Marcus 2010). In addition, there is a rather high state-dependency in terms of ownership status of wealth components, which facilitates the consideration of longitudinal information in the imputation process.

In many ways this work is a follow-up study to the evaluation study of single imputation methods for income panel data conducted by Watson and Starick (2011) with data from the Australian HILDA survey. They conclude their study with a few remarks: future research should test the performance of imputation methods under different assumptions concerning the non-response mechanism, an issue that we are trying to address in this study. Furthermore, they focused on single imputation methods and left it to other researchers to evaluate the performance of multiple imputation methods. Again, this is something we are tackling with this study. In our simulation study we compare six combinations of cross-sectional and longitudinal imputation strategies for German wealth panel data collected for the German Socio-economic Panel Study (SOEP) in 2002, 2007 and 2012. We create simulation data sets by setting observed data points to missing based on three separate non-response generating mechanisms. We examine the performance of imputation models assuming the mechanisms are missing at random (MAR) or the data suffers by differential non-response (DNR). We test the performance of multiple imputation by chained equations (MICE, named after one of the first popular implementations, see Royston, 2004). We test a univariate imputation procedure for panel data known as the row-and-column method introduced by Little and Su (1989). Additionally, we test a regression specification with correction for sample selection

¹ The drawbacks of case-wise deletion strategies have been well documented (Little & Rubin, 1987).

including a stochastic error term, which was the standard imputation method for the SOEP wealth data in survey waves 2002 and 2007.

The Paper is organized as follows: Section 2 gives an overview of wealth surveys and their imputation strategies and of item non-response in the SOEP wealth data, Section 3 describes how we generate simulation data sets with missing values from observed cases. Section 4 explains the evaluation set-up in detail and the criteria we are choosing to compare the imputation methods. In Section 5 we summarize the imputation methods and discuss their strengths and weaknesses. Section 6 details the performance of these methods using our simulated wealth data derived from the SOEP. Section 7 concludes.

2 Wealth Surveys and Incidence of Item Non-Response in SOEP

Wealth Data

Household panel surveys typically provide their users with imputed information. However, such surveys differ with respect to the imputation strategies applied to address item non-response and also in the way how available longitudinal information is incorporated. In the following we present panel surveys that collect wealth information, and their imputation strategies. Their consideration might give useful clues for the imputation of wealth data in this study.

The recently established Eurosystem Household Finance and Consumption Survey (HFCS) is a household wealth survey conducted in 15 euro area countries and organized by the European Central Bank (ECB) (see ECB, 2013a). This survey uses an iterative and sequential regression design for the imputation of missing data, similar to the sequential approach we evaluate in this paper (see section 4.2). The method used by the HFCS is adopted from similar surveys by the Federal Reserve Board and Banco de España (see Kennickel, 1991 1998, Barceló, 2006). The number of imputates provided by the HFCS is five, which seems to be the generally agreed on number of imputations provided with survey data.² In most of the participating countries the HFCS will be continued as a panel study (ECB, 2013b). However, the sequential approach the data providers are using has only been tried and tested in cross-sectional surveys thus far. We argue that the evaluation of multiple imputation strategies for longitudinal wealth data will increase in relevance in the future.

² The same number of imputates is also provided by e.g. the SCF, the SOEP, and SHARE.

The Survey of Health, Aging and Retirement in Europe (SHARE) is a cross-national panel survey including more than 85,000 individuals from 20 European countries aged 50 and older. SHARE also imputes data using a method that is similar to MICE (see Christelis, 2011).

The Household, Income and Labour Dynamics in Australia Survey (HILDA) is a household-based panel study which collects information about economic and subjective well-being, labour market dynamics and family dynamics in Australia (see Watson & Wooden, 2002). HILDA uses a combination of nearest neighbor regression imputation and the row-and-column imputation, depending on the availability of longitudinal information from other waves of the survey (Hayes & Watson, 2009).

The US panel study of income dynamics (PSID) is the longest running household panel survey, it started in 1968. The PSID asks about nine broad wealth categories; INR is imputed using a single hot-deck imputation technique, home equity is imputed using a simple carry-forward method (see PSID, 2011).

The German Socio-economic Panel Study (SOEP)—the survey used for our study—is a longitudinal representative survey collecting socio-economic information on private households in Germany (Wagner, Frick, & Schupp, 2007). In contrast to other wealth surveys that interview only one household representative, the SOEP collected wealth information separately for all household members (with age 17 or older) in 2002, 2007 and 2012. This survey strategy seems to be advantageous compared to collecting wealth information by one reference person per household only, given that accuracy and comparability to official statistics seem to perform better (Uhrig, Bryan, & Budd, 2012). One major drawback of this strategy is inconsistency on the household level. Given that asset values held by several household members can deviate from each other and may result in an even higher share of INR. The major disadvantage of surveys collecting the data solely interviewing one reference person is that the risk to overlook wealth, assets or debts of other household members increases. However, the methods we test in this evaluation study can be easily applied to wealth data collected at the household level and we do not expect the results to be significantly different in such a set-up.

The first wave of SOEP data was collected prior to the German reunification in 1984 with 12,245 respondents. The original sample was eventually supplemented by 10 additional samples to sustain a satisfactory number of observations and to control for panel effects. In 2002, an additional sample of high-income earners was implemented (2,671 individuals), which is particularly relevant for the representation of high net worth individuals in the sample given that income and wealth is rather highly correlated. In 2012, more than 21,000 individuals were interviewed.

The SOEP wealth module collects 10 different types assets and debts: value of owner-occupied and other property (and their respective mortgages), private insurances, building loan contracts, financial assets (such as savings accounts, bonds, shares), business assets, tangibles and consumer credits.

A filter question is asked whether a certain asset is held by the respondent, then the market value is collected and finally information about the personal share of property is requested (determining whether the interviewee is the sole owner or, if the asset is shared, the individual share).

For the imputation of the wealth data, there are three steps involved (for more information see Frick et al. 2007, 2010): Firstly, the filter imputation determines whether an individual has a certain asset type in his or her portfolio. These variables are imputed using rather simple logit regression models. Secondly, the metric values of the respective assets are imputed. And thirdly, a personal share is imputed again with a rather simple logit regression. In our simulation study we concentrate on the imputation of item non-response (INR) for the metric asset values.³

In table 1 we summarize the observed INR incidences for the SOEP wealth data 2002, 2007 and 2012 for the metric values. The respective share of INR varies between about zero for debts on other property and about 14 percent for private insurances.

³ (Partial) unit nonresponse and wave nonresponse—persons or households dropping out of the sample for a limited time or permanently—do not receive any imputation treatment in the person-level SOEP wealth data. Unit nonresponse generally is addressed by survey weighting procedures (see Kalton, 1986).

Table 1 | Item non-response rates in SOEP wealth questions

Wave	Type of wealth question	missing (metric) values*	share of missing values*	
2002 (n = 23,892)	gross wealth	home market value	1,104	4.60 %
		other property	453	1.90 %
		financial assets	1,822	7.63 %
		building-loan contract	(in 2002 together with private insurances)	
		private insurances	3,308	13.85 %
	gross debt	business assets	350	1.46 %
		tangible assets	592	2.48 %
		debts owner-occupied property	63	0.26 %
		debts other property	6	0.00 %
		consumer credits	366	1.53 %
2007 (n = 20,886)	gross wealth	home market value	1,093	5.23 %
		other property	364	1.74 %
		financial assets	1,931	9.25 %
		building-loan contract	921	4.41 %
		private insurances	2,781	13.32 %
	gross debt	business assets	290	1.39 %
		tangible assets	214	1.02 %
		debts owner-occupied property	179	0.86 %
		debts other property	40	0.19 %
		consumer credits	212	1.02 %
2012 (n = 18,361)	gross wealth	home market value	958	5.22 %
		other property	341	1.81 %
		financial assets	1,469	8.00 %
		building-loan contract	812	4.42 %
		private insurances	2,385	12.99 %
	gross debt	business assets	270	1.47 %
		tangible assets	196	1.07 %
		debts owner-occupied property	276	1.50 %
		debts other property	53	0.29 %
consumer credits	219	1.19 %		

Source: SOEP v29; (*) Note that the absolute number of missing metric values, as well as the share, is determined by the sample members who did report that they are holding a certain asset type and could not or refuse to provide a value, it excludes all members who did not report filter information, which has yet to be determined in a separate pre-value imputation. That is why for some variables with a low incidence (such as business assets) the filter information is missing for more individuals than the metric value.

3 Simulating Nonresponse

The first step in every imputation procedure that accounts for INR in a data set is to make an assumption concerning the nonresponse mechanism, which may be either explicitly formulated or implicitly derived from the imputation framework. The commonly used framework for missing data inference traces back to Rubin (1976), who differentiates the response mechanism for three assumptions: Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR). If the observation is assumed to be MCAR the probability of an observation being missing does not depend on any observed or unobserved variables. With MCAR, excluding all observations with missing values will yield unbiased estimators, but will also result in a loss of efficiency. Under MAR, given the observed data, the missing values do not depend on unobserved variables. That is, two units with the same observed values will share the same statistical behavior on other variables, whether observed or not. If neither of the two assumptions holds, the data is

assumed to be MNAR: the response status is dependent on the value of unobserved variables (e.g. the missing value itself) and cannot be accounted for by conditioning on observed variables.

The most commonly used assumption about the nonresponse mechanism is MAR. However, “as with other statistical assumptions, [...] the missing at random assumption may be a useful approximation even if it is believed to be false” Allison (1987, 77). Thus in the following we will focus on the evaluation of the imputation methods described in Section 4 only under MAR and two variants of MNAR.

We opt to focus on three components of the asset portfolio covered by the SOEP: home market value, financial assets and consumer credits. Home market value is easily the most important component in the average wealth portfolio in Germany. Financial assets are subject to both comparatively high non-response rates and rather high incidences. Additionally, regression models for the home market value tend to yield a good model fit, whereas models for financial assets tend to have a relatively poor model fit (Frick et al., 2007). This is equally true for both prediction models of the asset values and modelling the nonresponse mechanism itself. We chose consumer credits as the third component to cover in this study, because it exhibits rather low incidences and tends to fare mediocre as far as modelling is concerned; the reason is that the imputation cannot rely on a high number of sound covariates given that the SOEP does not collect additional information about this type of liability in comparison to other assets.

Since there still remains a large pool of fully observed observations after blanking out all INR cases, this turns out to be useful for the creation of simulation data sets. Depending on component and wave there are between 2291 and 8103 nonzero asset values (see the sum of ‘Number to be imputed’ and ‘Nonzero observations’ in table 1). Since it is not possible to compare imputed values with the true ones in our imputation set-up, we need to go one step back and create a simulation data set. Basically, we estimate a set of logit regression models for the non-response mechanism from all cases fully observed in any of the three waves of the SOEP wealth data.

Variables included in the non-response model are the employment status and the total personal income, the interview mode, a set of socio-demographic variables (e.g. gender, age, number of children, years of schooling, region) and a rather small set of supplemental economic indicators (e.g. financial support received). Additionally, a set of dummies indicate non-response in other wealth components in the same survey wave and a lagged dummy variable indicates non-response of the same variable in one of the other waves as state-dependency matters for INR in subsequent waves (Frick & Grabka, 2005). Those set of dummies covering the observed response behavior is among the most significant variables, when modelling the observed response behavior in the sample population.

Their incorporation requires that we do not blank out observed values in our simulation data sets based on a static prediction; we rather build a dynamic procedure that updates those predictions based on the response behavior in other waves and for the other two wealth components.

However, since the predicted probability that the value of a certain wealth component is highly dependent on whether the value has been observed in any of the two other waves, the share of observations in our simulation data sets with non-response in every wave was too high compared to the original dataset, as the information on the response status in other waves is the most important predictor. Therefore we added a small stochastic component to the predictions to incorporate uncertainty. After the addition of this random error terms the share of observations for which information from the other two waves is available for longitudinal imputation is approximately the same as in the original datasets.

Table 2 displays the McFadden R^2 for the non-response models under MAR, the number of observations with missing values and the number of nonzero observations for the simulation assets and waves. Note that the number to be imputed is fixed at around 10 percent of all valid nonzero observations, which is a rather high non-response incidence for home market value and consumer credits. The share of missing values for questions concerning the financial assets tends to be higher than 10 percent. However, since our performance criteria solely focus on the differences between imputed and observed data sets using only the respective imputed cases, this handicap does not have relevance in this study.

Table 2 | Descriptive statistics for observed and simulated data

INR assumption	Wave		McFadden R ²	Mean in Euro	Number to be imputed	Nonzero observations	Coefficient of Variation
OBSERVED	2002	Home market value	-	243,769	-	7075	0.731
		Financial assets	-	39,798	-	8103	3.209
		Consumer Credits	-	26,544	-	2088	4.792
	2007	Home market value	-	237,508	-	6775	0.762
		Financial assets	-	40,114	-	8377	3.651
		Consumer Credits	-	17,935	-	2978	2.850
	2012	Home market value	-	230,613	-	6164	0.726
		Financial assets	-	44,740	-	7377	2.901
		Consumer Credits	-	16,866	-	2552	4.911
MAR	2002	Home market value	0.595	225,724	707	6368	0.773
		Financial assets	0.410	44,921	810	7293	2.026
		Consumer Credits	0.524	26,475	208	1880	1.733
	2007	Home market value	0.518	214,858	677	6098	0.746
		Financial assets	0.391	54,026	837	7540	6.060
		Consumer Credits	0.618	16,191	297	2681	2.048
	2012	Home market value	0.540	202,057	637	5527	0.789
		Financial assets	0.406	59,015	737	6640	3.010
		Consumer Credits	0.597	18,689	255	2297	1.871
DNR I	2002	Home market value	-	204,609	716	6359	0.634
		Financial assets	-	15,762	808	7295	1.894
		Consumer Credits	-	10,168	176	1912	1.801
	2007	Home market value	-	190,218	692	6083	0.756
		Financial assets	-	11,242	809	7568	2.917
		Consumer Credits	-	6,190	301	2677	2.304
	2012	Home market value	-	195,064	636	5528	0.873
		Financial assets	-	11,287	773	6604	2.306
		Consumer Credits	-	6,682	256	2296	1.871
DNR II	2002	Home market value	-	283,085	760	6315	0.705
		Financial assets	-	73,853	805	7298	2.253
		Consumer Credits	-	39,505	209	1879	1.748
	2007	Home market value	-	284,654	637	6138	0.800
		Financial assets	-	75,950	858	7519	2.690
		Consumer Credits	-	41,856	309	2669	2.334
	2012	Home market value	-	301,754	626	5538	0.924
		Financial assets	-	84,956	763	6614	2.629
		Consumer Credits	-	36,835	261	2291	6.917

Source: SOEP v29, the number of observations to be imputed in the simulated data sets vary slightly around 10 percent of the nonzero observations in the observed data sets, as the exact number of missing values in each data set depends on a stochastic components under both MAR and MNAR.

However, as useful and necessary as MAR as an assumption for researchers to address item non-response is, to assume the (non-)response mechanism is fully explained once we conditioned on observed variables may putting things too simple. This is why we simulate two additional response mechanisms under the assumption of differential non-response: in two different set-ups we assume that the probability to provide the value of a certain asset depends on the value itself. The empirically observed relationship between nonresponse incidence and the corresponding values tends to be U-shaped, which is better documented for income questions than it is for wealth questions: In fact, Frick and Grabka (2005) state that the incidence for nonresponse of a component of the post-government income for the lowest and highest income deciles is between 28 and 60 percent higher than for the fifth and sixth income deciles. Additionally, characteristics that are

typically observed for low income and low wealth households, such as level of schooling and part time employment, have significant explanatory power in non-response models (Riphahn and Serfling, 2005). As Kennickell and Woodburn (1997) conclude with U.S. wealth data, the higher the household wealth is, the higher the probability that the household refuses to participate.⁴

Under the assumption that wealth components share a similar non-response behavior, we assume in the DNR1 data sets that the probability that a value is missing is the higher, the lower the true value is (i.e. differential non-response at the bottom of the distribution). In the DNR2 data sets, we assume the contrary, the higher the true value of the wealth the higher is the probability that the value is missing. Table 2 compares the effects on the mean and the coefficient of variation of the respective simulation data sets. Consequently, the means for the observations to be imputed in the DNR1 data sets are substantially lower, whereas in the DNR2 data sets they are substantially higher than in the data sets containing all observed cases.

4 Evaluation Criteria

For the choice of evaluation criteria, we follow a different path from the evaluation framework laid down by Watson and Starick (2011) and focus on a set of 8 different instead of 11 suggested criteria applied by the authors. The main applications of wealth data—not only of SOEP—are divided in three sections. (1) Cross-sectional analyses focus on point estimates, trend and distributional analyses. (2) Inequality measurement focuses on the computation of the GINI coefficients and other inequality indices. (3) Longitudinal analyses focus on wealth mobility. (1) and (2) are rather closely related and should be adequately replicated by the imputation procedure. (3) is an additional focus, which is tackled in a separate evaluation. Hence, we divided the evaluation criteria into two subsets, to account for the comparatively higher importance of wave-specific trend and inequality analyses (six criteria in section 4.1) compared to rare analyses that specifically make use of the panel structure of the data (two additional longitudinal criteria in section 4.2). Ultimately, an ideal imputation model would account for cross-sectional, longitudinal *and* inequality accuracy.

⁴ Vermeulen (2014) gives a comprehensive overview of the potential effects of differential non-response for high-net-worth-individuals on the measurement of inequality in the European HFCS survey data.

4.1 Wave-Specific Evaluation Criteria

Finding suitable evaluation criteria for multiple imputation is challenging. Most criteria applied by Watson and Starick (2011) are not applicable to the task at hand, as they would be heavily biased in favor of a replication of the observed value; for instance, an evaluation of the correlation between observed and imputed value does neglect the fact, that it is not the goal of multiple imputation to create a valid value for an individual missing item, but rather create a valid data set that takes the uncertainty of the imputation procedure into account. Hence, multiple imputation is best understood as simulating values for valid inference. In this study, we chose to evaluate trend, distributional and inequality accuracy jointly in a set of six evaluation criteria that take the overall data set into account instead of the replications of single values.

Chambers (2001) notes the imputation results should reproduce the lower order moments of the distribution of the true values. Given that we can directly compare the lower order moments between imputed and observed data sets, we chose to include the **absolute relative difference in means (1)** for the assessment of trend accuracy and the **absolute difference in the coefficient of variation (2)** as an indicator of distributional and inequality accuracy.

$$CR(1) = \left| \frac{(\bar{y} - \hat{y})}{\hat{y}} \right|$$

$$CR(2) = \left| \frac{\sigma}{\bar{y}} - \frac{\hat{\sigma}}{\hat{y}} \right|$$

Additionally, distributional accuracy is achieved when the distributional properties of the original data set is replicated by the imputed data sets. The **Kolmogorov-Smirnov distance (3)** is the higher the more the two tested empirical distributions of the imputed and the true values deviate from each other. Thus, the smaller the Kolmogorov-Smirnov distance is, the more accurate the imputation method.

$$d_{KS} = \max_j \left(\left| \frac{1}{n} \sum_{i=1}^n I(y_i \leq x_j) - \frac{1}{n} \sum_{i=1}^n I(\hat{y}_i \leq x_j) \right| \right)$$

For the assessment of inequality we include three additional criteria. The **Gini coefficient (4)** is especially sensitive against changes in the center of the distribution. The **mean log deviation (5)** is sensitive for shifts at the bottom of the distribution. Those two criteria are complemented by an inequality measure for the top tail of the distribution, by using the **99/50 ratio of percentiles (6)**.⁵

⁵ This indicator is not responsive to outliers—a relevant phenomenon in wealth analyses—compared to e.g. the half squared coefficient of variation (HSCV).

4.2 Additional Longitudinal Evaluation Criteria

We apply two additional evaluation criteria that help to examine the effects of the imputation on wealth mobility. The first criterion assesses the distributional accuracy of wealth mobility between waves for specific components and includes all observations with a positive value for the specific wealth type in two waves simultaneously. Here, wealth mobility is defined by the change in wealth decile group membership in 2002 vs. 2007, 2007 vs. 2012 and 2002 vs. 2012. A standard Chi-square test for fit of the distributions is performed where the imputed cell frequencies are the observed ones and the expected cell frequencies are the true cell frequencies.

$$\chi^2 = \sum_{j=1}^{10} \sum_{i=1}^{10} \frac{(\hat{n}_{ij} - n_{ij})^2}{n_{ij}}$$

Thus, the higher the **Chi-square test statistic (7)** the worse the imputation method can replicate the observed mobility for the wealth component in consideration.

The second longitudinal criterion is the **cross-wave correlation (8)** for each wealth type separately: before and after the imputation procedure the differences of the correlations between each wealth type are compared and should be close to zero. The higher the deviation from zero the worse the performance of the imputation method.⁶

$$r_{y_1 y_2} - r_{\hat{y}_1 \hat{y}_2} = \left| \frac{\sum_{i=1}^n (y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (y_{i1} - \bar{y}_1)^2 \sum_{i=1}^n (y_{i2} - \bar{y}_2)^2}} - \frac{\sum_{i=1}^n (\hat{y}_{i1} - \bar{y}_1)(\hat{y}_{i2} - \bar{y}_2)}{\sqrt{\sum_{i=1}^n (\hat{y}_{i1} - \bar{y}_1)^2 \sum_{i=1}^n (\hat{y}_{i2} - \bar{y}_2)^2}} \right|$$

⁶ For comparison's sake we need to mention that we opt to not include four criteria applied by Watson and Starick (2011) that we find do not add another dimension to the evaluation at hand and, thus, are redundant. This includes the preservation of skewness and kurtosis, since the replication of the shape of the distribution is covered by the Kolmogorov-Smirnow distance (3). Furthermore, unlike Watson and Starick (2011) we do not include Pearson correlations between two wealth types. There is not enough covariation for this criterion to be applied for the asset types we choose for this study.

5 Imputation Methods

The imputation methods which can be considered in our simulation study are limited by the fact that we are interested to use multiple imputation techniques. We have to rule out all single imputation techniques beforehand. This includes for example all carryover methods which use valid values observed in the last or next wave of the survey (and variations thereof, which have been applied in the PSID for home equity). This also includes, more generally, all imputation methods without a stochastic component. The methods we choose to examine are commonly used by other important wealth surveys, as we already referenced in the second chapter.

We also refrain from considering (longitudinal) hotdeck imputation given that Watson and Starick (2011, 711) already present evidence in a simulation study that the hotdeck imputation method does “not perform particularly well on either cross-sectional or longitudinal accuracy”.

5.1 Multiple Imputation by Chained Equations (MICE)

MICE is an iterative and sequential regression approach that grew popular among researchers, because it demands very little technical preparation and is easy to use. We present the basic set-up for imputations using chained equations in this chapter, but for more detailed information we refer to van Buuren, Boshuizen and Knook (1999), Royston (2004), and van Buuren, Brand, Groothuis-Oudshoorn and Rubin (2006), among others. Multiple imputation by chained equations (MICE) is not an imputation model by itself, it is rather the expectation that by sequentially imputing the variables using separate univariate imputation models there will be convergence between the imputed variables after a certain number of iterations. For each prediction equation all but the variable for which missing values ought to be imputed are included, that is, each prediction equation exhibits a fully conditional specification. It is necessary for the chained equations to be set up as an iterative process, because the estimated parameters of the model are possibly dependent on the imputed values. Formally, we have p wealth components Y_1, Y_2, \dots, Y_p and a set of predictors (without missing values) Z , then for iterations $n = 0, 1, \dots, N$, and with ϕ_j as the corresponding model parameters with uniform prior probability distribution, the missing values are drawn from

$$\begin{aligned}
Y_1^{(n+1)} &\sim g_1(Y_1|Y_2^{(n)}, \dots, Y_p^{(n)}, Z, \phi_1) \\
Y_2^{(n+1)} &\sim g_2(Y_2|Y_1^{(n+1)}, Y_3^{(n)}, \dots, Y_p^{(n)}, Z, \phi_2) \\
&\dots \\
Y_p^{(n+1)} &\sim g_p(Y_p|Y_1^{(n+1)}, Y_2^{(n+1)}, \dots, Y_{p-1}^{(n+1)}, Z, \phi_p)
\end{aligned} \tag{1}$$

until convergence at $n = N$ is achieved. That is, in iteration $n + 1$ the dependent variables of each imputation model $g_j(\cdot)$ are updated with the corresponding imputed values of the last iteration n (or the ongoing iteration, if the dependent variable already has been imputed). One of the main advantages is that the univariate imputation models $g_j(\cdot)$ may be chosen separately for each imputation variable, which is also why in spite of a theoretical justification for MICE, it is widely used by researchers and practitioners. We did not make use of this specific feature at the project at hand, as all wealth variables exhibit similar statistical and distributional characteristics. However, we choose an adjusted set of additional independent variables Z_j for each imputation variable Y_j . In line with the experiences of other countries and surveys for the imputation of wealth data, the additional independent variables Z_j we choose are a set of (1) covariates determining the non-response (variables of the non-response model under the MAR assumption mentioned in section 4.1.), (2) covariates that are considered good predictors for the variable we want to impute (3) economic variables that are possibly related to the outcome variable (according to economic theory) and (4) variables that are good predictors of the covariates included in the rest the groups of variables. However, the last group is especially important in the first iterations and the more association between the imputation variables is expected. Nonetheless, we follow those guidelines for the independent variables in the prediction equations and refer to Barceló (2006) for an overview on the reasoning behind the extensiveness of the set covariates and some examples. To give an example why we adjusted the set of independent variables for each imputation variables: e.g. regional information tends to have significant explanatory power for the imputation models of real estate but do not contribute to the estimated models for most of the remaining wealth components.

We specified the imputation models $g_j(\cdot)$ in (1) using predictive mean matching (PMM) to account for the restricted range of the imputation variables and to circumvent the assumption that the normality of the underlying models holds true. Predictive mean matching (PMM) was introduced by Little (1988) and is a nearest-neighbor matching technique used in imputation models to replace the outcome of the imputation model for every missing value (a linear prediction) with an observed value. The set of observed values from which the imputed value is randomly drawn consists of (non-

missing) values derived from the nearest neighbors which are closest to the linear prediction. Thus, the distribution of the observed values will be preserved for the imputed values.

5.2 Regression with Heckman Correction for Sample Selection

For the first two waves of wealth information in the SOEP, the researchers opted for a regression design with Heckman correction for sample selection for the imputation of the missing asset values (Frick et al. 2007, 2010). The first step involved a cross-sectional imputation of missing values for 2002. These data were then used for a longitudinal imputation of the 2007 data using the lagged wealth data from 2002 as covariates. The third step was a re-imputation of 2002 wealth data using the now-completed longitudinal information from 2007, and starting a cycle of regression models with longitudinal info until convergence between 2002 and 2007 was achieved. The stochastic component in each step, which is necessary to generate multiple imputates, was added through the assignment of randomly drawn residuals derived from the respective regression models. As for this study, we decided to include this already deployed approach in our simulation to compare its performance with other multiple imputation methods.

With the 2012 wealth data and three available waves, the pool of available longitudinal information grew considerably. We decide to add the regression models for 2012 after convergence between 2002 and 2007 has been achieved, with 2007 now serving as the base year. Consequently, longitudinal information from the survey wave 2007 is used for the imputation of missing values in 2002 and 2012 alike.

The variables included in those models are similar to the set of covariates used in the MICE approach (see Section 4.1). However, this regression approach is not sequentially adding updated imputed values from other wealth types; hence the models, predictions and imputed values are calculated isolated, the prediction equation does not include the metric values of the other wealth types.⁷

5.3 Row-and-Column Imputation Technique

Little and Su (1989) proposed the row-and-column imputation technique (RC) as a procedure for item nonresponse adjustment in panel surveys. It takes advantage of available cross-sectional as well

⁷ There are a few exceptions: The regression model for home value (other property values) additionally includes the home debt (other property debt). The imputations for both these values are generated in an iterative process in itself, since both values have very high explanatory power in the respective models.

as individual longitudinal information. It combines data available from the entire panel duration for every unit (row) and cross-sectional trend information (column) and adds a residual derived from a nearest neighbor matching, thereby attaching a stochastic component to an otherwise deterministic approach.

Since we have three waves of wealth data, the column effects (for any wealth asset) are given by

$$c_t = \frac{(3 * \bar{y}_t)}{\sum_k \bar{y}_k} \quad (2)$$

and are calculated for each wave separately. \bar{y}_t is the sample mean wealth asset for $t = 2002, 2007, 2012$. The row effects are given by

$$r_i = \frac{1}{m_i} * \sum_j \frac{y_{it}}{c_j} \quad (3)$$

and are calculated for each member of the sample. y_{it} is the value of the wealth asset for individual i in wave t . m_i is the number of recorded waves in which the asset value of individual i has been observed.

Originally, the row-and-column-method was designed as a single imputation method. However, the last step—assigning the residual term from the nearest neighbor—may be modified in such a way that for every individual unit and wave multiple imputed values can be derived. After sorting the units by their row effects r_i , the residual effect of the nearest complete unit l in year j is used to calculate the imputed value for unit i :

$$\hat{y}_{it} = r_i * c_t * \frac{\text{residual term}}{r_l * c_t} \quad (4)$$

\hat{y}_{it} is the single imputed value using the residual effect from the nearest neighbor l . To generate multiple imputations we need only two additional steps. Instead of only assigning the residual of the nearest neighbor in (4), we assign the residuals of the k nearest neighbors. Then terms (2) and (3) are identical for every computation and n residual terms are used to generate k imputed values for every unit i and every year t . Since there is a tradeoff between the number of imputations and the distance to the “farthest” nearest neighbor, we reasoned that the generally agreed on number of five imputations would present a reasonable balance (see e.g. the HFCS, other SOEP-variables, the Survey of Consumer Finances (SCF)). However, this decision is merely based on our expectations and has not been subject to an empirical analysis. Also it is noteworthy, that the residual terms of the five nearest-neighbors have been randomly assigned to imputed values independently for every unit i in order to avoid any systematic differences of imputation accuracy in the five imputation data sets.

5.4 Row-and-Column Imputation with Age Classes

When using the row-and column imputation the donor of the residual term (and the distance between donor and recipient) in (4) is solely depending on the sorting of the units by their row effects r_i . Additionally, the trend component (2) is calculated using the complete sample. At the same time, as Watson and Starick (2011) state, recipients and the respective donors should have similar characteristics, and those characteristics should be associated with the variable being imputed. They introduce an addition to the basic row-and-column imputation; the method is extended to take into account basic characteristics of the donors and recipients. For a comparison between the standard row-and-column imputation and an imputation with age classes (RCA) (see figure 2) we match donors and recipients within longitudinal imputation classes defined by the following age classes (at the time, the survey was conducted) in the respective wave: 17–19, 20–24, 25–34, 35–44, 45–54, 55–64, 65 and older. Thereby it is guaranteed that donors will share their residual with recipients from the same age range. The column term (2) will be calculated using observations from the respective age classes.

An restriction of the Row-and-Column imputation is that it cannot be applied if no longitudinal information on the person level is available, thus we need a fallback method in case only cross-sectional information is at hand (e.g. the first wave of a respondent, or a specific wealth component is collected for the first time). As for the evaluation, we need a set-up that determines the superior combination of basic and fallback imputation methods simultaneously (see table 3). The results of the evaluation should provide answers to several questions: (1) If a row-and-column imputation is used for observations that have valid information in other waves, does the addition of age classes improve the performance when compared to the standard row-and-column imputation? (2) Which combination of basic and fallback methods yields the best results? Basic imputation method means the technique that is used for observations with missing values and values from other waves of that same individual have been observed. Fallback imputation method means that for an observation with missing values only cross-sectional information and variables are available and, therefore, only either of the two model based approaches can be applied. Hence, in addition to the combinations using model based and row-and-column imputations, we test the performance of using a multiple imputation by chained equations as both basic and fallback method (MICE), and we proceed similarly with the regression with Heckman correction (REG).

Table 3 | Basic and fallback imputation methods, and evaluation set-up

BASIC (for observations with missing values, information from other waves is available)	Standard Row-and-column imputation (Little & Su 1989)		Row-and-column imputation (Little & Su 1989) using age classes		Multiple imputation by chained equations	Regression model with Heckmann correction for sample selection
FALLBACK (for some observations with missing values, only cross-sectional information and variables are available)	Multiple imputation by chained equations	Regression model with Heckmann correction for sample selection	Multiple imputation by chained equations	Regression model with Heckmann correction for sample selection		
acronym used in chapter 5	MICE-RC	REG-RC	MICE-RCA	REG-RCA	MICE	REG

6 Results

As we illustrated in table 3, we compare the performance of the six combinations of prevalent imputation methods using the eight evaluation criteria we discussed in section 4. As we wanted to compare the performance of the methods on a metric scale, we refrain from any ranking of the results. Second, we favor the property that the punishment for large deviations is larger than for smaller deviations, which should depend on the overall variance of the outcomes considering the individual evaluation criteria. That means, if the overall variance is small, outliers will be punished harder, and deviations that are close to each other should be punished similarly. Again, this is a property that is not fulfilled by any ranking of the results. It is, however, fulfilled, if we choose a distance measure that shows the distance between a well-defined optimum and the respective values calculated with imputed data. The optimum is simple to define, as all criteria are either calculated in a way that zero is representing no deviations from the original data or may be transformed to have this respective property. As for the distance measure, using the Euclidian distance would either require a normative decision on a weighting matrix or, alternatively, all criteria would contribute similarly (after normalizing). In order to avoid normative weighting we choose the Mahalanobis distance measure, as it additionally accounts for the observed covariance structure (Mahalanobis, 1936), and thereby is removing any redundancy in our evaluation criteria.

Our evaluation shows the distance between the ideal imputation (all values are zero for all criteria) and the deviation of the imputed values from this ideal point after using the respective imputation method (all tables in section 6). Furthermore, this evaluation set-up allows us to compare the

distances directly and interpret them on a metric scale, as the respective outcomes for the different methods are independent from each other (but depending on the overall variation and covariation of the evaluation criteria).

As already mentioned, we show the results for the three wealth items, the three years, and the three assumed nonresponse mechanisms separately and compare the outcomes for the imputation methods. The evaluation criteria (1) – (6) are used for the trend, distributional and inequality evaluations. The longitudinal criteria (7) and (8) are additional criteria, which can solely be computed using the joint results of two waves (2002/07, 2007/12 and 2002/12) as reported in section 6.2.

6.1 Evaluation of Trend, Distributional and Inequality Accuracy

If we would have solely considered the home market value in this study (table 4), we would conclude that all combinations including the RC imputation yield better results than the pure REG and MICE imputations: Only taking into account the average distances for the trend evaluation reveals that in most cases the MICE and REG imputations perform worse than the combinations with the RC imputation with and without age classes. Looking at the performance for all single waves, in all but two cases the addition of the RC technique as basic imputation improves the performance of MICE. Combining REG with the RC imputation on the other hand does not regularly improve the results. What is even more surprising, even though the combination of MICE and RC technique seems to perform best overall, the pure MICE approach rarely performs better than the pure REG approach. A possible explanation for these findings is that the home market values tend to be an asset type with a rather high state-dependency. The RC approach as univariate imputation technique, which solely considers future and past observed values and an overall trend effect, is closer to the trend and inequality estimates based on the observed data sets than both model-based approaches that may incorporate the uncertainty of the imputation procedure. Note that these outcomes are basically independent of the non-response mechanism that is assumed.

Table 4 | Overall performance of home market value imputation methods

Wave-Specific Evaluation				
	2002	2007	2012	Overall Average Distance
Assumption: Missing at Random				
REG	4.60	3.61	4.49	4.24
REG-RC	3.74	4.32	4.22	4.09
REG-RCA	4.91	3.95	4.29	4.38
MICE	5.13	4.40	5.21	4.91
MICE-RC	3.68	3.48	3.83	3.66
MICE-RCA	3.78	3.48	4.12	3.79
Assumption: Differential Non-Response 1				
REG	4.42	5.10	4.75	4.76
REG-RC	4.83	4.80	4.43	4.69
REG-RCA	4.30	3.61	4.20	4.04
MICE	4.90	4.65	4.74	4.76
MICE-RC	4.52	4.05	4.47	4.35
MICE-RCA	4.13	2.78	4.89	3.93
Assumption: Differential Non-Response 2				
REG	4.24	4.76	4.67	4.56
REG-RC	3.96	3.20	4.50	3.89
REG-RCA	4.06	3.41	4.55	4.01
MICE	5.35	4.95	4.47	4.92
MICE-RC	3.32	3.10	3.88	3.43
MICE-RCA	3.07	3.34	3.69	3.37

Bold figures indicate the smallest average distance among the six imputation variants.

Generally, financial assets exhibit less state-dependency than home market values and regression models for both the imputation of the metric values and the nonresponse mechanism are mediocre compared to other asset types (table 5). Thus, there is comparatively more uncertainty to consider by the imputation method, and the lag or lead variables have, in theory, considerably less explanatory power. However, if the missing mechanism is MAR, combining MICE with the RC method, again, yields the best results. If the missing mechanism is differential non-response at the bottom of the distribution, MICE-RCA seems yield the best results as well. Only if differential non-response at the top is assumed, it is equally viable to choose between any RC method including age classes. Interestingly, including age classes oftentimes improves the results for the RC technique. One possible explanation might be that the value of the assets under consideration regularly increase or decrease their value depending on the age of the asset holder, thereby including age classes might reduce the uncertainty of the imputation process.

Interestingly, for the evaluation criteria that are considered in this study and for financial assets, it seems to be more viable to choose a pure REG approach over a pure MICE approach. Combining REG and RCA on the other hand barely improves the results except under DNR2. However, it is notable

that all combinations of MICE with the RC method again regularly perform better than both pure model based approach.

Table 5 | Overall performance of financial assets imputation methods

Wave-Specific Evaluation				
	2002	2007	2012	Overall Average Distance
Assumption: Missing at Random				
REG	3.94	4.21	4.38	4.17
REG-RC	4.19	4.34	4.63	4.39
REG-RCA	4.12	4.67	4.57	4.45
MICE	5.08	5.07	4.77	4.97
MICE-RC	3.39	4.14	3.06	3.53
MICE-RCA	3.35	3.78	3.04	3.39
Assumption: Differential Non-Response 1				
REG	4.67	5.25	3.73	4.55
REG-RC	4.49	4.51	4.66	4.55
REG-RCA	4.50	4.36	4.61	4.49
MICE	5.17	5.06	4.66	4.96
MICE-RC	4.36	3.64	4.52	4.18
MICE-RCA	4.56	3.83	4.23	4.21
Assumption: Differential Non-Response 2				
REG	5.01	5.19	5.59	5.26
REG-RC	4.43	4.12	4.38	4.31
REG-RCA	4.31	4.14	4.58	4.34
MICE	5.59	5.04	4.90	5.18
MICE-RC	4.08	4.48	5.18	4.58
MICE-RCA	4.28	3.33	4.61	4.07

Bold figures indicate the smallest average distance among the six imputation variants.

Consumer credits have the lowest state-dependency of the three wealth types we consider in this study. Note that the SOEP wealth data is collected in five-year intervals and credit periods for consumer credits are typically shorter. Following the same argumentation we already laid out for home market values and financial assets, we expect that the RC imputation performs rather weak. The results of the evaluation prove us mostly wrong. As shown in table 6, both RC methods perform better if MAR or DNR1 is assumed. One possible explanation is that even if the overall state-dependency is much lower for consumer credits, the state-dependency at the bottom of the distribution may still be considerably high and the RC imputation might still yield more accurate imputed data sets in this case. If DNR2 is assumed, the results are less clear: RC or RCA does improve the REG imputations slightly, RCA improves the MICE imputation slightly and yields the best results overall. Independently of the non-response mechanism assumed, it seems to be wise to include age classes in the RC technique.

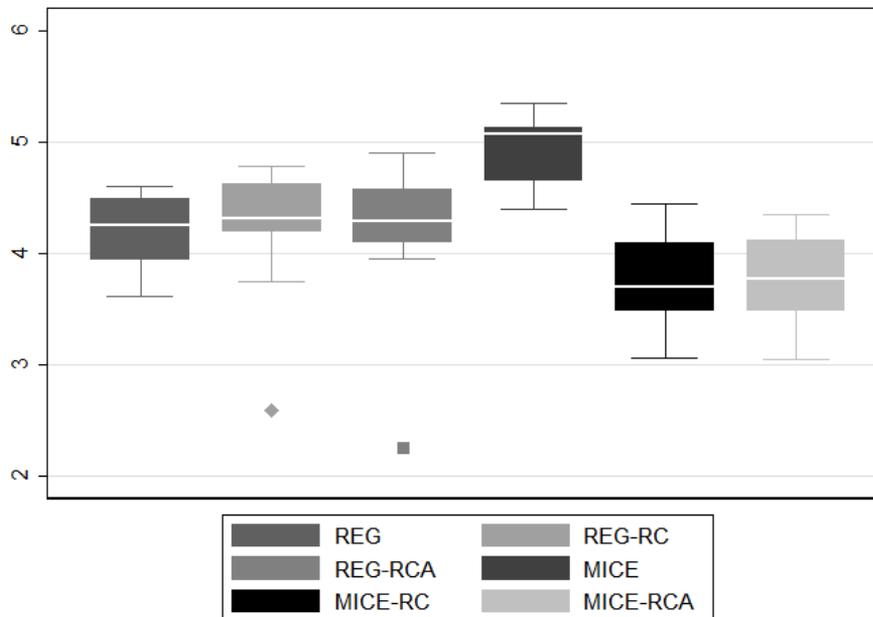
Table 5 | Overall performance of consumer credits imputation methods

Wave-Specific Evaluation				
	2002	2007	2012	Overall Average Distance
Assumption: Missing at Random				
REG	4.25	4.51	3.83	4.20
REG-RC	4.79	4.62	2.59	4.00
REG-RCA	4.09	4.31	2.25	3.55
MICE	5.35	4.65	4.63	4.88
MICE-RC	4.44	3.70	4.10	4.08
MICE-RCA	4.34	3.48	4.24	4.02
Assumption: Differential Non-Response 1				
REG	4.97	4.36	4.48	4.60
REG-RC	5.52	3.90	3.44	4.29
REG-RCA	4.39	3.95	3.84	4.06
MICE	5.30	5.26	4.97	5.18
MICE-RC	4.55	4.50	4.38	4.48
MICE-RCA	4.22	4.38	4.51	4.37
Assumption: Differential Non-Response 2				
REG	4.96	4.56	5.77	5.10
REG-RC	4.77	5.16	4.51	4.81
REG-RCA	4.85	4.86	4.39	4.70
MICE	5.07	4.85	4.63	4.85
MICE-RC	5.09	4.89	4.80	4.92
MICE-RCA	4.41	4.71	4.74	4.62

Bold figures indicate the smallest average distance among the six imputation variants.

Comparing the distributions of the distances to the optimal imputations separately for the MAR assumption for all three waves and all assets jointly, confirms the conclusions we draw above (figure 1).

Including the univariate RC imputation does improve the performance of MICE considerably and significantly. The distance between the optimal imputation and MICE versus both MICE-RC and MICE-RCA is roughly 1.2 units higher, the respective means and standard errors are shown in figure 1 together with the boxplots of the distributions. Considering the performance of REG versus REG-RC and REG-RCA the differences are miniscule. Similar figures for DNR1 and DNR2 are buried in appendix A.



	Mean	Std. err.		Mean	Std. err.
REG	4.20	0.11	MICE	4.92	0.11
REG-RC	4.16	0.22	MICE-RC	3.76	0.14
REG-RCA	4.13	0.26	MICE-RCA	3.73	0.15

Figure 1 | Boxplots for the distances to optimal imputations by imputation methods under Missing at Random (MAR)

Additionally, we observe that the incorporation of age classes in the RC imputation oftentimes improves the overall imputation results. Watson and Starick (2011) report an advantage for the performance of the RC imputation with age classes for the imputation of income items as well. One possible explanation, why identify a similar advantage, is that there is a regular trend of increase and spend-down of asset values over the life cycle.

6.2 Evaluation of Wealth Mobility

As for the two additional longitudinal criteria, which focus on the changes in the observed mobility structures before and after imputations, the overall average distances include all pair-wise comparisons (2002/2007, 2007/2012, and 2002/2012) and are presented in Table 7. We expected that the univariate imputations using RC imputations would overestimate the state-dependency for

the wealth assets and undermine the actually observed mobility structures. This expectation gets confirmed to a certain extent.

Table 7 | Average performance on longitudinal evaluation criteria, all assets

Longitudinal Evaluation				
	Home Market Value	Financial Assets	Consumer Credits	Overall Average Distance
Assumption: Missing at Random				
REG	2.05	1.27	0.76	1.36
RC-REG	2.19	1.44	0.91	1.51
RCA-REG	2.46	1.77	0.44	1.56
MICE	0.19	0.89	0.91	0.66
RC-MICE	2.17	1.73	0.59	1.50
RCA-MICE	2.14	1.76	0.62	1.51
Assumption: : Differential Non-Response 1				
REG	0.82	2.98	0.98	1.59
RC-REG	2.07	3.07	1.11	2.08
RCA-REG	2.07	3.02	1.02	2.04
MICE	1.14	1.46	0.77	1.12
RC-MICE	1.16	1.17	0.49	0.94
RCA-MICE	1.16	1.17	0.56	0.96
Assumption: : Differential Non-Response 2				
REG	2.88	1.71	0.67	1.75
RC-REG	3.17	1.47	1.04	1.89
RCA-REG	1.41	1.41	0.94	1.25
MICE	0.29	1.44	0.60	0.78
RC-MICE	1.00	1.48	1.85	1.44
RCA-MICE	1.10	1.48	1.60	1.39

Bold figures indicate the smallest average distance among the six imputation variants.

Under MAR the pure MICE approach seems to perform better than the pure REG approach and all combinations with the RC method (at least for home market value and financial assets). This is to be expected, as the mobility seems to be severely reduced, once the only included variable is the lag or lead variable of the respective variable that is to be imputed. What is more surprising is that the REG approach performs considerably worse than MICE. One possible statistical explanation could be that the regression set-up is not taking into account one source of uncertainty, which the MICE procedure does take into account: the drawing of the respective model parameters. The only stochastic component in the REG approach is the drawing of a residual from the observed residuals, whereas MICE imputes values after drawing of the respective model parameters. Here, REG might underestimate the uncertainty of the imputation procedure and produce too less variation in the imputed values, thereby as well reducing mobility.

For differential non-response and especially consumer credits the results are less clear. Generally MICE seems to reproduce the observed mobility structures better than REG, in many cases the combination of MICE and the RC imputations yield satisfying results too. We conclude that (1) a researcher interested in mobility structures would probably prefer the model based MICE approach to an univariate imputation procedure such as the RC method, and (2) even though REG yields imputed values using model prediction equations as well, the REG imputation performs considerably worse than the MICE approach.

7 Conclusion

In an assessment of the performance of several imputation methods for longitudinal wealth data we use a set of eight evaluation criteria and three assumptions for the non-response mechanism. The overall result does not yield that a single imputation method performs consistently better for all wealth types in a cross-sectional and longitudinal analysis. We compare the row-and-column imputation (with or without age classes) for observations with available longitudinal data with two methods that rely on the prediction equations of regression models. In our analyses of the performance of the imputation methods we identified several effects the researcher has to consider for studies using multiple imputation and imputed data.

As for the trend and inequality evaluation, if the missing data are truly missing at random (MAR) the differences between the regression based approach with Heckman and correction for sample selection and the inclusion of the row-and-column method are generally rather small. However, for all three assets we consider, the combination of MICE and row-and-column imputation is at least among the best performing methods. Unexpectedly, this holds true independently of the level of state-dependency prevalent in the items. If the missing data are missing not at random and instead are the result of differential non-response (DNR1 and DNR2) the combination of the univariate row-and-column imputation with one of the two model-based approaches does improve the performance without any exception in our evaluation study. This is the core outcome of this study: If the missing at random assumption is violated, the row-and-column imputation technique yields less biased overall imputation results for trend and inequality estimates. We like to stretch that—based on this study and our experience with data imputation—this conclusion holds only true for variables that are highly skewed (such as assets, net worth or income variables). The imputation technique itself—and thus an improvement of the performance—is applicable to panel data only.

Furthermore, we find that adding age classes to the standard row-and-column imputation as introduced by Little and Su (1989) does slightly improve the performance based on our criteria and the input data.

As for the wealth mobility criteria, the conclusions are less clear. Generally, MICE seems to reproduce the observed mobility structures better than the regression approach, in many cases the combinations of MICE and the row-and-column imputations yield satisfying results, too. However, it is noticeable for some assets and nonresponse assumptions that the mobility is reduced, once the row-and-column imputation is applied.

One thing that remains to be addressed is that we refrained from including partial unit nonresponse (PUNR) in this simulation, e.g. individuals within households that choose not to respond, whereas the rest of the household did. The reason is that analyses with the SOEP wealth data focus on the individual level observation and PUNR observations would only affect household wealth estimators. However, we do not expect the results to be significantly different, had we considered PUNR observations. Potential extensions to this study could be the inclusion of additional wealth types, examining the effects of imputation methods on the total net worth and the aggregate net worth, and additional imputation methods we did not consider for now.

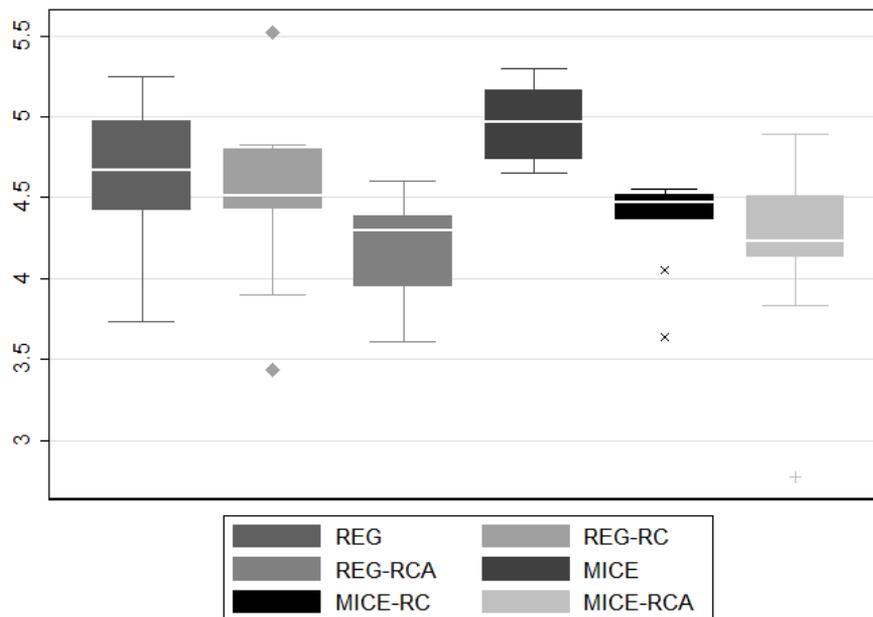
8 References

- Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. Clogg (Ed.), *Sociological Methodology*, 17, 71–103.
- Barceló, C. (2006). Imputation of the 2002 wave of the Spanish survey of household finances (EFF). Banco de España, *Documentos Ocasionales* 0603. DOI: <http://dx.doi.org/10.2139/ssrn.901584>.
- Chambers, R. L. (2001). Evaluation Criteria for Statistical Editing and Imputation. National Statistics Methodological Series, 28, University of Southampton.
- Christelis, D. (2011). Imputation of Missing Data in Waves 1 and 2 of SHARE. DOI: <http://dx.doi.org/10.2139/ssrn.1788248>.
- European Central Bank (ECB). (2013a). The Eurosystem Household Finance and Consumption Survey. Results from the first wave. Statistics Paper Series 2. Frankfurt/Main: ECB. Available at: <http://www.ecb.europa.eu/pub/scientific/stats/html/index.en.html> (accessed June 2015)
- European Central Bank (ECB). (2013b). The Eurosystem Household Finance and Consumption Survey. Methodological report for the first wave. Statistics Paper Series 1. Frankfurt/Main: ECB. Available at: <http://www.ecb.europa.eu/pub/scientific/stats/html/index.en.html> (accessed June 2015)
- Frick, J. R., Grabka, M.M., & Marcus, J. (2007). Editing and Multiple Imputation of Item-Non-Response in the 2002 Wealth Module of the German Socio-Economic Panel (SOEP). DIW Berlin Data Documentation 18. Available at: <http://hdl.handle.net/10419/86162> (accessed June 2015)
- Frick, J.R., & Grabka, M. M. (2005). Item nonresponse on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility. *Allgemeines Statistisches Archiv*, 89, 49-61.
- Frick, J.R., Grabka, M.M., & Marcus, J. (2010). Editing und multiple Imputation der Vermögensinformation 2002 und 2007 im SOEP. DIW Berlin Data Documentation 51. Available at: <http://hdl.handle.net/10419/86163>.
- Hayes, C., & Watson, N. (2009). HILDA Imputation Methods. Hilda Project Technical Paper Series 2/09. Melbourne Institute of Applied Economic and Social Research. Available at: https://www.melbourneinstitute.com/hilda-research/Survey_Methods_and_Data.html (accessed June 2015)

- Kalton, G. (1986). Handling Wave Nonresponse in Panel Surveys. *Journal of Official Statistics*, 3, 303–314.
- Kennickel, A. (1991). Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation. Federal Reserve Working Paper Series, Washington DC.
- Kennickel, A. (1998). Multiple Imputation in the Survey of Consumer Finances. Federal Reserve Working Paper Series, Washington DC.
- Kennickell, A., & Woodburn, R. L. (1997). Consistent Weight Design for the 1989, 1992, and 1995 SCFs, and the Distribution of Wealth. Federal Reserve Board, Survey of Consumer Finances Working Papers.
- Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6, 287–296.
- Little, R.J.A., & Su, H.L. 1989. Item nonresponse in Panel Surveys. In *Panel Surveys* edited by D. Kasprzyk, G.J. Duncan, G. Kalton, & Singh, M. P., 400-425, New York: Wiley.
- Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2, 49-55.
- Panel Study of Income Dynamics (PSID). (2011). *Documentation for the 2007 PSID Supplemental Wealth File. Release 2: March 2011*. Available at: <http://psidonline.isr.umich.edu/Data/Documentation/wlth2007.pdf> (accessed June 2015).
- Riphahn, R., & Serfling, O. (2005). Item non-response in income and wealth questions. *Empirical Economics*, 30, 521-538.
- Royston, P. (2004). Multiple Imputation of missing values. *Stata Journal*, 4, 227-241.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, 4, 87–94.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, Wiley.
- Schenker, N., & Taylor, J. M. G. (1996). Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis*, 22, 425–446.

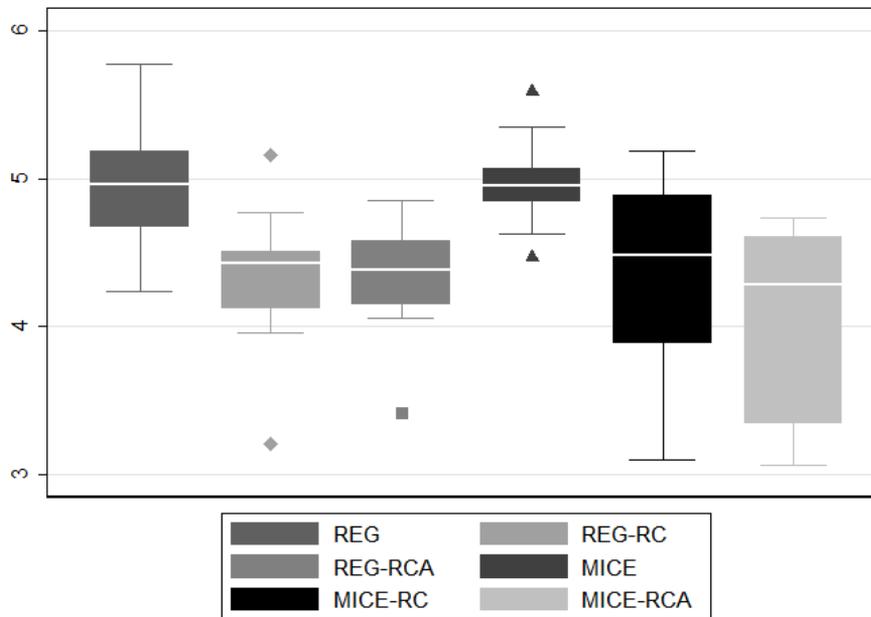
- Socio-Economic Panel (SOEP) (2013), data for years 1984-2012, version 29, SOEP, 2013, doi:10.5684/soep.v29.
- Uhrig, N., Bryan, M., & Budd, S. (2012). "UKHLS Innovation Panel Household Wealth Questions: Preliminary Analysis." Understanding Society Working Paper Series No. 2012 – 01, January 2012. Available at: <https://www.iser.essex.ac.uk/publications/working-papers/understanding-society/2012-01> (accessed June 2015).
- van Buuren, S., Boshuizen, H. C., & Knook, D.L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18, 681–694.
- van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D.B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76, 1049–1064.
- Vermeulen, P. (2014). How fat is the top tail of the wealth distribution? Working Paper Series 1692, European Central Bank.
- Wagner, G.G., Frick, J.R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch, Journal of Applied Social Science Studies*, 127, 161-191.
- Watson, N., & Starick, C. (2011). Evaluation of Alternative Income Imputation Methods for a Longitudinal Survey. *Journal of Official Statistics*, 27, 693-715.
- Watson, N., & Mark, W. (2002). The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Survey Methodology. HILDA Project technical papers series No. 1/02, May 2002 (Revised October 2002). Available at: https://www.melbourneinstitute.com/hilda-research/Survey_Methods_and_Data.html (accessed June 2014).

Appendix



	Mean	Std. err.		Mean	Std. err.
REG	4.64	0.15	MICE	4.97	0.08
REG-RC	4.51	0.20	MICE-RC	4.33	0.10
REG-RCA	4.19	0.11	MICE-RCA	4.17	0.20

Figure A1 | Boxplots for the distances to optimal imputations by imputation methods under Differential Non-Response 1 (DNR1)



	Mean	Std. err.		Mean	Std. err.
REG	4.97	0.16	MICE	4.98	0.11
REG-RC	4.34	0.18	MICE-RC	4.31	0.25
REG-RCA	4.35	0.15	MICE-RCA	4.02	0.22

Figure A2 | Boxplots for the distances to optimal imputations by imputation methods under Differential Non-Response 2 (DNR2)