USC Dornsife Center for Economic and Social Research

USC Schaeffer Leonard D. Schaeffer Center for Health Policy & Economics

The Effect of Using Different Imputation Methods for Economic Variables in Aging Surveys

Jinkook Lee, Erik Meijer, Drystan Phillips

Paper No: 2015-019

CESR-SCHAEFFER WORKING PAPER SERIES

The effect of using different imputation methods for economic variables in aging surveys

Jinkook Lee, Erik Meijer[*], Drystan Phillips

University of Southern California, Center for Economic and Social Research

August 18, 2015

Abstract

We study the sensitivity of analyses on income and wealth to the methods of imputation for missing data. We do so by implementing a conditional hot-deck imputation, based on the method used by the English Longitudinal Study of Ageing, in the Health and Retirement Study. We compare marginal and joint distributions and regression models estimated from data using these imputations to the same statistics estimated from the RAND HRS data, which use a much more elaborate imputation method, and to estimates from data using an intermediate method. Although many results are qualitatively or quantitatively similar, there are also notable differences.

1. Introduction

The accurate measurement of variables such as income and wealth is essential for many economic studies. Unfortunately, most data sets economists work with are taken from surveys in which these variables are measured less than perfectly. The extent and consequences of measurement error in surveys has therefore received considerable attention in the literature; see, for example, Duncan and Hill (1985), Bound and Krueger (1991), Bollinger (1998), and the overview in Bound, Brown, and Mathiowetz (2001). A book-length treatment of econometric modeling with measurement error can be found in Wansbeek and Meijer (2000). Missing data can be viewed as an extreme form of inaccurate measurement. Some forms of missing data, such as sample selectivity (e.g., Heckman, 1979) and attrition in panel data (e.g., Fitzgerald, Gottschalk, & Moffitt, 1998) have received considerable attention, but there has been limited research on the practical consequences of missing data on the results of empirical economic studies. The theory of missing data patterns and how to correctly analyze data sets that have missing values can be found in the statistical literature, for example, Little and Rubin (2002). Cameron and Trivedi (2005, chapter 27) briefly discuss this literature and give a simulated example.

The missing data problem can be alleviated in several ways. The easiest one for users of the data is using *imputations* of the missing values, preferably provided by the agency that released the data. The goal of imputation is to replace the missing values by random draws from their conditional distribution given the observed data. When properly done, this preserves the relations among variables and corrects for selection bias under the assumption that data are *missing at random* (MAR; Rubin, 1976; also called *selection on observables*; e.g., Fitzgerald, Gottschalk, & Moffitt, 1998). This assumption allows missingness to depend on the observed

[*]Corresponding author; erik.meijer@usc.edu.

values, but not on the missing values themselves. It is unlikely to be exactly satisfied for the types of missing data that we study here, but results tend to be quite robust to small-to-moderate violations of the assumption and in any case simpler methods will generally be more biased (Graham 2009). When a broad range of covariates is used in modeling the conditional distribution for the imputations, results using imputed data should not suffer appreciably from selection bias. For users of the data, imputation is convenient because standard complete-cases analysis methods can be used to analyze the data.

Any imputation method is based on an explicit or implicit model. Different imputation methods are based on different models, or different variations of the same model (e.g., different covariates). Because it is generally unknown what the "correct" model is, there can be some discussion about what the preferred imputation method is, and what the consequences are of using different imputation methods. In the economic literature, this has received some attention, specifically regarding the imputation of wages and salary income in the Current Population Survey (CPS), with mixed conclusions. David et al. (1986) compared the CPS hot deck imputation with a regression-based imputation and found only very small differences between the results using both. Moreover, they compared their results with results obtained using matched IRS earnings and concluded that the hot deck "does not underestimate income aggregates to any serious extent," whereas Lillard, Smith, and Welch (1986), studying the mechanics of the hot deck procedure in great detail, conclude that "Clearly, the census procedure severely understates income in certain occupations. Because it is based on the apparently invalid assumption that income does not affect reporting propensities, it most likely understates average incomes as well." Bishop, Formby, and Thistle (2003) compare two versions of the CPS method that were both released for 1988: the version that used the hot deck method that was used pre-1988, which in some cases overwrote reported information with imputations, and a revised method, which does not overwrite the reports. They also compared this with a regression-based imputation method. They concluded from these comparisons that "failure to correct for nonparticipation is a serious and continuing problem in the current CPS processing system." Similar to the arguments in Lillard et al. (1986), Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) argue that the CPS hot deck imputation method tends to impute too many values in the middle of the distribution, which leads to an attenuation bias they call *match bias*, which is similar to the attenuation bias when covariates are measured with error, although the match bias they study occurs with imputations in the *dependent* variable.

Hoynes, Hurd, and Chand (1998, section 8.5) compare the effects of different imputation methods on the marginal distribution of wealth in the Asset and Health Dynamics among the Oldest Old (AHEAD) data, which studies individuals age 70 and older and their spouses. Since 1998, this study has been merged into the Health and Retirement Study (HRS), and their preferred imputation method is a predecessor of the imputation methods currently used in the HRS. They find relatively small differences in marginal distributions between the different versions of the imputation methods they use, with the exception of an unconditional hot deck that ignores the information from the unfolding brackets (see below for a description of these), which has different means and medians—sometimes in opposite directions. The magnitudes of these differences are between 5 and 20 percent.

In this paper, we study the effects of using different imputation methods in modern surveys on aging, specifically the HRS. This is particularly relevant for a number of reasons. First, these data contain imputed variables that have been used in a growing number of studies in labor economics (e.g., Gustman & Steinmeier, 2005; van der Klaauw & Wolpin, 2008), health economics (e.g., Michaud et al., 2011; Hurd et al., 2013), and especially the intersection of the two (e.g., Benítez-Silva et al., 1999; French, 2005), as well as many other fields, such as health, sociology, and epidemiology. Hence, understanding the quality of these data is of interest in itself.

Second, a unique aspect of these surveys on aging is that they are part of a worldwide concerted effort of collecting comparable panel data, with the intention of learning about the role of institutions on health, economic outcomes, and well-being of the elderly and near-elderly (see, e.g., Banks & Smith, 2012; Erosa, Fuster, & Kambourov, 2012; Banks et al., 2015). However, the surveys in different countries use different imputation methods (see section 3). Hence, the results from this paper will help understanding whether any cross-country differences found in comparative analyses of aging surveys reflect actual population differences or may (partly) be due to differences in imputation methods.

Third, these aging surveys have been at the forefront of innovations in questionnaire design aimed at improving the measurement of economic variables. The prime example is the use of *unfolding brackets*, which are follow-up questions after a failure to report a continuous amount, in order to obtain at least a limited range in which the value is situated (Juster & Smith, 1997). This reduces the imputation uncertainty, and thus the potential for large consequences of using different imputation methods. Another example is the integration of questions for income and wealth and matching the periodicity over which income questions are asked to the typical way such income is received (Hurd, Juster, & Smith, 2003), which greatly improved the quality of asset income (which increased substantially) and Social Security income, which may have consequences for the imputations as well.

Fourth, we study not only earnings, but also asset income (dividends) and wealth, as well as the correlations between these income and wealth components. Thus, the results from our study are relevant for a much broader range of economic variables than the studies about earnings imputations in the CPS.

We study the effect of the imputation method by implementing a version of the method that the English Longitudinal Study of Ageing (ELSA) uses in the data of the Health and Retirement Study (HRS) and then comparing distributions and regression results obtained with these imputations to those using the imputations that are in the public release data of the HRS. This means we are comparing one of the simplest imputation methods with one of the most complex ones, on the same data. We also implement an intermediate, regression-based imputation method and compare this with the other two methods.

In section 2, we discuss the theoretical rationale for imputation, highlighting a number of aspects that may affect the quality of the imputations and the expected consequences of potential choices. In section 3, we describe the data and the details of the imputation methods that we will compare. Section 4 presents the results, and section 5 concludes.

2. Theoretical background

Suppose that the collected data for an individual i consist of two sets, y_i and x_i and individuals are drawn from a population by simple random sampling. However, some respondents do not report y_i and only report x_i. Let r_i denote whether y_i is observed ($r_i = 1$) or missing ($r_i = 0$). Suppose the goal of an analysis is to estimate a parameter θ, say (possibly vector-valued), which is a function of the cumulative distribution function $F_{(y,x)}(\bullet, \bullet)$ of (y_i, x_i). Usually, θ will be a smooth function of $F_{(y,x)}$ and consistent estimation of $F_{(y,x)}$ then implies consistent estimation of θ. We can now write

$$F_{(y,x)}(a,b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f_{(y,x)}(y,x)\,dy\,dx = \int_{-\infty}^{b} F_{y|x}(a|x)\,f_x(x)dx, \tag{1}$$

where uppercase F denotes cumulative distribution functions and lowercase f denotes probability density functions, and the subscripts denote whether it is a joint, conditional, or marginal distribution or density. This notation implicitly assumes that the variables are continuously distributed. With discrete distributions, lowercase f denotes a probability mass function, and integral signs should be interpreted as summations, and analogously for mixed discrete-continuous variables. From (1), it follows that we can estimate $F_{(y,x)}$ consistently if we can estimate $F_{y|x}$ consistently. The latter can be written as

$$\begin{aligned} F_{y|x}(a|x) &= \Pr(y \le a|x) \\ &= \Pr(y \le a|x, r = 1)\Pr(r = 1|x) + \Pr(y \le a|x, r = 0)\Pr(r = 0|x). \end{aligned}$$

Imputations are typically based on the MAR assumption, which implies $\Pr(y \le a|x, r = 0) = \Pr(y \le a|x, r = 1)$, that is, the conditional distribution of y given x does not depend on whether or not y is missing. Then, we can estimate this conditional distribution from the observed data and use (1), and we do not need to estimate a model for the missingness process itself and for $\Pr(y \le a|x, r = 0)$. The way this is implemented is by generating a random draw y_i^* of y from $F_{y|x}$ for every observation i in which y is missing and then estimating θ as if y_i^* was an observed value. Under the MAR assumption and some regularity conditions, this gives consistent estimators. Note that it is essential that the imputations are *draws* from the conditional distribution (or posterior distribution), not *estimates* of the unknown values y_i (e.g., conditional mean or posterior mean), because the latter would underestimate the true variation.

When missingness is *completely at random* (MCAR), we also have that $\Pr(r = 1|x) = \Pr(r = 1)$, that is, missingness does not depend on x either. Then the distribution of the observed data is the same as the distribution we are interested in, $F_{(y,x)}(a, b \mid r = 1) = F_{(y,x)}(a, b)$, so analyses using only the observations without missings (called *complete cases analyses* or *listwise deletion*) give consistent estimators and the only effect of the missingness is a reduction of sample size and hence precision. Then imputation will often still be useful, to increase precision. In general, however, missingness will not be completely at random and then the distribution of the observed data will be different from the distribution of interest, even if the MAR assumption is satisfied, and thus θ will often not be consistently estimated without correction. (Note that what is called y and x here depends on missingness, and may not coincide with what would be dependent and exogenous variables in the model of ultimate interest.) As we have seen, under the MAR assumption, imputation is a way to correct this.

When there are multiple patterns of missingness in the data, the definition of r generalizes such that r becomes an integer variable with a different value for each pattern of missingness, and the definition of which variables are in y and x depends on the missingness pattern. These are technical details that do not change the theory materially.

In theory, (1) implies that *all* variables that are observed should be included in x. However, a typical survey has thousands of variables, perhaps even more than observations, so including all of them in an imputation model is impossible in practice. Thus, the imputation model needs to balance the risk of overfitting with the risk of model misspecification. If only a subset w of x is included, this implies the assumption that the elements v, say, of x that are not included in w are conditionally independent of y, given w.

Similarly, (1) implies that the imputation model should estimate the *joint* distribution of all elements of y conditional on x, and in particular the model and the imputations should preserve the correlations between the elements of y after controlling for x. Again, this may be a daunting task, and the imputations in many surveys impute each variable separately. There are exceptions, however, like the iterative procedures of Raghunathan et al. (2001) and van Buuren et al. (2006), which cycle through a set of potentially missing variables and impute them in turn, using the others as covariates. These techniques are occasionally applied in practice, for example in the Survey of Health, Ageing and Retirement in Europe (SHARE; Börsch-Supan & Jürges, 2005; Christelis, 2011).

If the elements of y are imputed independently, the correlation between the imputed variables is lower than in the population. Including more covariates in the imputation model (i.e., in w) reduces this bias, because a higher fraction of the variation in each y variable is accounted for by the systematic part of the imputation model, leaving a residual term with lower variance, and thus a smaller potential contribution of the residual to the correlation between the y variables. Similarly, more covariates in w imply that the conditional independence assumption regarding the omitted variables v is less problematic.

Often, there is a huge number of potential covariates for the imputation models. Some of these may be highly correlated themselves, in which case using only one of them (or of a set of correlated variables) does not affect the quality of the imputations and any analyses based on them as much. For example, suppose the "true " model is $y = a + bw + cv + e$ and w and v are related by $v = p + qw + u$. If v is omitted from the imputation model, the estimated model is $y = (a + cp) + (b + cq)w + (e + cu)$. If w and v are highly correlated, that is, the R-square of the regression of v on w is high (q is large relative to the variance of u), the omission of v from the imputation model does not affect the quality of the imputations greatly. Note that the goal of imputation is to reflect the joint distribution of the observed variables, not to estimate causal coefficients consistently, so the "omitted variables bias" is not a problem in itself. However, if the correlation between w and v is low but v is strongly correlated with y, its omission does affect the imputations considerably. This analysis shows why it is advisable to include a small number of variables from each of a broad set of domains that are thought to be related to the variable that needs to be imputed and to the probability of a missing response. It also illustrates why imputation under the MAR assumption with a broad set of covariates greatly reduces any biases even if the assumption is not strictly true, because the MAR assumption is another form of a conditional independence assumption, namely conditional independence of y and r conditional on the covariates used in the imputation model.

In addition to the selection of covariates to include in the imputation model, imputation models can differ in the functional forms they use or distributional assumptions they make. As with the choice of covariates, this involves a trade-off between flexibility and generality on the one hand, which reduce bias, and parsimony on the other hand, which avoids overfitting and imprecise estimates that may lead to outliers in the imputations, especially for observations where the covariates are in the tails of their distributions.

Most surveys provide only one imputed value for each missing value, Thus, to the user, the released dataset looks like a dataset without any missing responses (at least for the imputed variables), which makes it easy to analyze with any statistical or econometric method designed for complete data. However, treating the imputations as if they were reported values underestimates the standard errors, because the variability introduced by the imputation uncertainty is not taken into account. In some cases, it is possible in principle to obtain standard errors that take the imputation uncertainty into account (e.g., Abadie & Imbens, 2012), but this requires detailed knowledge of the imputation method, and often the ability to replicate the imputation method (e.g., in bootstraps), which is impractical and sometimes even impossible. An easier solution for the user is to provide *multiple imputations*, which some surveys do (e.g., SHARE and the Korean Longitudinal Study of Aging [KLoSA], KLI, 2007). The advantage of multiple imputations is that standard errors and other forms of statistical inference can easily take imputation uncertainty into account. See, for example, Rubin (1987, 1996) and Schenker and Welsh (1988) for a detailed account of the theory and practice of multiple imputation. The drawbacks of multiple imputation are larger storage requirements and computational burden than with a single imputation, and more cumbersome analyses, although the latter has been much improved by the availability of standard multiple imputation facilities in widely used statistical packages, such as Stata and SAS. However, it is our impression that users without a strong statistical background tend to use only one of the imputations. In this paper, we focus on the estimates and not so much on the standard errors, and we do not compare multiple imputation with single imputation. Tentative results from other experiments in the kind of data we are studying here suggest that ignoring the imputation uncertainty leads to a downward bias of less than 10 percent in the standard errors, which is generally not a first-order problem.

3. Data and imputation methods

Many countries around the world face aging populations, which pose challenges in health care, labor markets, public pensions, and other areas. To understand these challenges, and the lives of the elderly and near-elderly in general, a large number of aging studies have been initiated, starting with the Health and Retirement Study (HRS) in the U.S. in 1992 (Juster & Suzman, 1995; NIA, 2007). Aging studies are ongoing in North and Latin America, Europe, and East and South Asia. These studies are panel studies, which repeatedly interview the same respondents over time, typically once every two years. A unique characteristic of these studies is the concerted harmonization efforts to enable cross-country studies: questionnaires are made similar to the greatest extent possible (accommodating country-specific issues such as different pension systems) and the investigators leading these studies regularly meet to achieve ex-ante harmonization. Such coordination allows researchers to perform cross-country comparisons in order to learn about the effects of institutions on labor force participation, physical and mental health, economic and psychological well-being, and other outcomes of interest. The Gateway to Global Aging Data website (https://g2aging.org/) indexes these surveys, allowing comparisons of questionnaires and searching for topics across studies, helps create datasets for cross-country comparisons, and provides a set of longitudinal data files of harmonized variables ready to be used for cross-country analysis.

A large part of the aging surveys is devoted to measuring respondent and household income and wealth. The questionnaires ask about detailed components of income, assets, and liabilities. Unfortunately, as in almost all surveys, respondents sometimes fail to answer these questions, either not knowing the exact amount of an income or wealth component or refusing to answer the question. This is a problem, because a small to moderate amount of missing data on each component may mean a large loss of information regarding total household income or wealth, arguably the most used income and wealth variables. Moreover, missing data may not be completely random: individuals who receive a certain kind of income or own a certain asset are more likely to be unsure about its value than those who do not have it. Therefore, a variable is more likely to be missing when the value is high. This results in biases when only complete cases are analyzed. For example, the mean of income for complete cases will tend to be lower than the mean in the population of interest.

Many of the aging studies therefore include imputed data for missing values in income, wealth, and some other variables. Plans are underway to impute missing data for the remaining aging studies as well. However, the imputation methods used by the different studies vary widely. The imputations as done by RAND for the HRS (Moldoff et al., 2014; Hurd et al., 2015) are based on a system of regression models with a broad set of covariates. By contrast, the English Longitudinal Study of Ageing (ELSA; Marmot et al., 2003), uses a simple hot-deck method based on cells, which for most variables are defined by the cross-classification of only two demographic variables and the known range within which the variable to be imputed should fall (Oldfield, 2014). In contrast with HRS and ELSA, the Survey of Health, Ageing and Retirement in Europe (SHARE; Börsch-Supan & Jürges, 2005) accounts for the correlations between income and wealth variables (conditional on covariates) by iteratively imputing from sequential regression models, which include the other income and wealth variables as regressors (Christelis, 2011).

3.1 The HRS data

The HRS is a large-scale multidisciplinary panel survey of individuals over the age of 50 and their spouses of any age. The HRS is the primary source of information about older individuals in the U.S. and as such plays a key role in scientific studies of this age group and in the evaluation of public policies affecting them. Among the main interests of economists and policy makers is the economic well-being of this age group, and in particular their income and wealth. As indicated above, the HRS introduced several methodological innovations in economic measurement, and the quality of its economic data is therefore highly regarded.

If the target individual is unable to answer the questionnaire (because of physical or mental health problems), the HRS interviews a *proxy* respondent instead, typically the spouse or other family member, who answers the question instead for the target individual to the extent possible and relevant (e.g., cognitive tests are necessarily skipped).

For couples, the income and wealth questions are answered by the *financial respondent*, for both spouses or partners. In a small fraction of the households in each wave, there is no financial respondent (e.g., the designated financial respondents dies before being interviewed but after the spouse has already been interviewed), in which case the whole income and wealth sections are missing.

For most income and wealth categories, the HRS first asks respondents whether they receive a type of income (e.g., wages or salary income) or own a type of asset (e.g., stocks). In the following, we refer to both of these as "ownership. " If the answer is yes, the HRS then asks the amount of income or the value of the asset. If the respondent is unable or unwilling to give an amount or value, the HRS then asks the respondent a sequence of *unfolding bracket* questions, such as, “Is it more than \$50,000, less than \$50,000, or about \$50,000?” Depending on the respondent’s answer, the HRS then asks one or more analogous follow-up questions with higher or lower amounts. The amounts (bracket thresholds), like the \$50,000 in the example, differ between questions and sometimes across waves. In later waves, including the wave we study here, the amount first presented is randomized. Results from this question sequence will indicate one of the following: (i) the household does not receive the income or own the asset; (ii) a continuous value of the income or asset; (iii) a range of values based on the respondent’s answers to bracket questions; (iv) missing (unknown) information on whether the household receives the income or owns the asset. In case (iii), we can further distinguish between completed or incomplete unfolding bracket sequences, or between an unbounded (open) range of the form “more than \$X,” a bounded (closed) range of the form “between \$X and \$Y,” and an “about” value ("about \$X"). In case (iv) we can distinguish between a specific don't know or refuse answer to this question and the failure to interview a financial respondent.

For this study, we focus on wave 9 (2008) of the HRS. RAND introduced two major improvements in the wealth variables in the 2013 release (version M of the RAND HRS): incorporation of corrections from the asset-verification section and including information from adjacent waves in the imputation models (cross-wave imputation). Replicating the asset-verification corrections in the hot-deck-imputed data would complicate our study greatly, while not adding additional insights on comparisons between the methods, especially as the HRS is the only survey that includes an asset verification section, and thus any conclusions about this would not generalize to other aging data sets. Therefore, we primarily use the last version of the RAND imputations before these changes were implemented (version L; St.Clair et al., 2011) to compare the different imputation methods.

We selected two income components and two asset components. The income components are income from wages and salary, which is a respondent-level variable, and dividend income, which is a couple-level variable. The asset components are amount of money in checking and savings accounts and value of direct stock and mutual fund holdings, that is, holdings outside Individual Retirement Accounts (IRAs) and defined-contribution pension accounts. These asset components are both couple-level variables. By selecting these four variables, we obtain information about both income and asset variables, and about correlations between income variables, between asset variables, and between income and asset variables. Because ELSA does not impute economic variables for institutional interviews, we also set nursing home interviews aside and only compare imputations for residential interviews.

Table 1 presents the response patterns. For three out of four variables, more than two thirds of the responses are “No income or asset”, and another 10 to 25 percent give a continuous value of the amount. The exception is checking and savings accounts, where 14% report not having these and about 60% give a continuous value. Thus, for the vast majority of the observations, we have an exact answer that does not require imputation. Also, imputations of “about” values are equal to the given number (e.g., if the respondent answers “about \$20,000,” then \$20,000 is imputed), so imputations of these are always the same and one could argue that these are not true imputations. However, they are treated as imputations in the RAND HRS data and in this paper (Kim & Hong, 2012, stochastically impute such cases). Among the cases that do need nontrivial imputation, “DK ownership” plus “No financial respondent” are a minority whereas for the other cases the respondent has indicated receiving the income or owning the asset. So most observations requiring imputation do receive the income or own the asset. This illustrates that missing data is not random: households not owning a component are less likely to have a missing value than the households who do own it. While not surprising, this illustrates the potential role of imputation in correcting for selective nonresponse bias.

3.2 RAND's imputation method for the HRS

For responses other than “no ownership” or a continuous value, RAND imputes a continuous value. This is done sequentially, according to the amount of information provided by the respondent (Moldoff et al., 2014).

(1) If ownership is unknown, it is imputed by a draw from a binary logit model. This model is estimated using all observations in which ownership is known (reported), with the set of covariates described below. An exception is when the number of observations in which the number of respondents who indicate owning the asset or income component is too low (less than 50) to estimate the logit model reliably. Then, ownership is imputed using just the unconditional fraction. This does not happen for the variables in this study.

(2) If the household owns the asset or income component (reported or imputed), but neither a continuous value nor a complete bracket has been given, a complete bracket is imputed by a draw from an ordered logit model. This draw is conditional on the information provided. For example, if the response results in an incomplete bracket that is the combination of two complete brackets, the draw is restricted to these two brackets, and the probabilities used are the conditional probabilities of being in each of the two brackets, given the covariates but also given that the observation must be in one of the two brackets. This is achieved by setting the probabilities of the inconsistent brackets to zero and scaling the probabilities of the consistent brackets such that they sum to 1. This model is estimated using all observations in which a complete bracket has been reported, with the set of covariates described below. An exception is when the number of observations in which a complete bracket was reported is too low (less than 50) to estimate the ordered logit model reliably. Then, the bracket is imputed using just the unconditional fractions. This also does not happen for the variables in this study.

(3) If a complete bracket with an open range has been reported or imputed, a continuous value is imputed by a draw from a lognormal censored regression (tobit) model. Again, this draw is consistent with the information provided, so this is a draw from a lognormal distribution truncated at the cutoff point of the top bracket. The estimation sample for this model includes all continuous reports (but not “no ownership"), but it censors the lowest 25 percent to the 25th percentile. The reason for this is that the aim of this model is to estimate the right tail of the distribution well. The censoring makes the results less sensitive to possible distributional misspecification of the left tail, while still keeping the responses to help in estimating the coefficients of the covariates more precisely. The set of covariates is again the same set described below. An exception is when the number of observations in which a continuous value was reported is too low to estimate the tobit model reliably (less than 50 in version L; less than 100 for the cross-wave imputations in version M and later). Then, the censored lognormal distribution is estimated without covariates and a value is drawn from this unconditional distribution (but still truncated at the top bracket cutoff). Again, because we study relatively common asset and income components here, this does not happen for the variables in this study.

(4) If a complete bracket with a closed range has been reported or imputed, a continuous value is imputed as the reported value of the *nearest neighbor* household: A linear regression model is estimated for the inverse hyperbolic sine[1] of the reported continuous values. For each missing observation, the donor pool consists of all observations with a *reported* continuous value within the closed range in which we need to impute. The nearest neighbor is the household in the donor pool with the closest *predicted* value from the estimated regression model. The imputed value for the target household is the *reported* value from the nearest neighbor. An exception is when the number of observations in which a continuous value was reported is too low (less than 50) to estimate the regression model reliably. Then, the value is imputed using a conditional hot-deck. That is, a donor record is randomly picked from the reported values in the same range. Again, this does not happen for the variables in this study.

In all these models, the covariates are the same. They consist of the first 10 principal components of a set of about 30 explanatory variables. The set of explanatory variables is slightly different depending on whether an income or asset component is imputed and on whether there is a financial respondent in the household (if not, the number of available variables is smaller). The explanatory variables for both income and asset components include demographics, education, health, cognition, and expectations about leaving bequests, and some interactions of these. The set of explanatory variables for the income components adds indicators of labor force status, whereas the set for asset components adds a few (imputed, if necessary) income variables. The cross-wave imputations in version M and later add a few longitudinal covariates in the models, in addition to the principal components: dummies for ownership of the same asset/income in the previous and next wave, the inverse hyperbolic sines of the values in these two waves, and dummies for marital status changes between the current wave and the two adjacent waves.

There are a few further exceptions and special cases for rare situations in which there are not enough observations to estimate a model or not enough donors. These do not affect the variables in this study. See St.Clair et al. (2011) and Hurd et al. (2015) for more details of the imputation methods and their rationale.

3.3 The ELSA imputation method

The English Longitudinal Study of Ageing (ELSA; Marmot et al., 2003) is the sister study of the HRS in England. ELSA was designed in close cooperation with key investigators affiliated with the HRS, to make the study not only useful and important for England, but to allow cross-country comparisons as well. Banks and Smith (2012) give an overview of publications based on international comparisons, many of which include a comparison of ELSA with the HRS.

The ELSA income and assets questions are similar to the HRS questions as described above, and thus lead to similar types of response cases. However, ELSA imputes data differently. Whereas the RAND method imputes each step in the sequence separately, ELSA imputes

[1] The inverse hyperbolic sine (asinh) is a transformation that is almost the same as a scaled log transformation for positive values not close to zero, but is defined and monotonic across the whole real line.

continuous values in a single step. ELSA's imputation method is described in detail in Oldfield (2014).

Imputations are done with a conditional hot-deck method based on a few variables. This is similar to the imputation method employed in the Current Population Survey (e.g., Abadie & Imbens, 2012). For each observation and each variable, a donor pool is selected with the same values on a few categorical (or discretized) variables and whose continuous reported values (including zeros for no ownership, when applicable) are consistent with the information from the missing observation. Thus, if the missing observation is accompanied by a complete, incomplete, or no bracket sequence, the donor pool is restricted to reported values within the closed or open range indicated by the bracket sequence. The imputed value for the missing observation is then the reported value for a randomly selected observation from the donor pool.

For couple-level (or *benefit-unit*-level in ELSA terminology) variables such as dividend income, checking and savings accounts, and stocks and mutual fund holdings, the variables that determine the donor pool (in addition to the bracket information) are benefit-unit type (couple, single man, single woman) and age (<50, 50 to state pension age, state pension age to 75, 75+). For different-sex couples, the age of the man is used, whereas for single-sex couples, the age of the oldest member is used. For individual-level variables such as wage and salary income, the variables that determine the donor pool (in addition to the bracket information) are age (in the same bands as for the couple-level variables) and gender. For a few other variables, alternative or additional variables are used to determine the donor pool.

ELSA gives couples the opportunity to answer some individual-level income variables separately or jointly (by the financial respondent), depending on whether they keep their finances separate while ELSA requires each respondent to answer other individual-level income variables on their own behalf. Sometimes, it happens that one member of a couple that keeps finances separate is a respondent and answers questions about his or her finances, whereas the other member is a non-respondent, leading to missings on half of these income components for the household, which does not happen in the HRS. Another difference between the HRS and ELSA is that ELSA asks some detailed questions about individual earnings while providing a summary measure at the benefit-unit level. The ELSA imputation methods have special rules for such cases. Furthermore, ELSA does not impute economic variables for institutional interviews (i.e., interviews in nursing homes) and a few other situations.

3.4 Our implementation and a third method

The HRS and ELSA imputation methods are very different. Given that ELSA was designed to be comparable with the HRS and is used frequently for comparisons between England and the U.S., it is important to assess to what extent any differences may be the result of using a different imputation method, rather than the result of differences between the English and U.S. populations. This will also be informative to users of other surveys that provide similar types of imputations. Furthermore, these results allow providers of other surveys that currently do not have imputations to make a more informed decision about what types of imputations, if any, they should provide. This is especially relevant for aging surveys in other countries, in particular the Japanese Study of Aging and Retirement (JSTAR; Ichimura et al., 2009), the China Health And Retirement Longitudinal Study (CHARLS; Zhao et al., 2013), and the Longitudinal Aging Study in India (LASI; Arokiasamy et al., 2012).

As stated in section 2, imputations should in principle use the conditional distribution of the variable to be imputed conditional on *all* observed variables (which differs between households). In practice, however, this is impractical, undesirable, and often impossible, because these data sets have huge numbers of variables, and this would lead to overfitting and inability to estimate all the coefficients in the models. Nevertheless, RAND's imputations are guided by this principle and start with a broad set of explanatory variables, which is then reduced by principal components analysis. In contrast, ELSA uses only two explanatory variables for most imputation models. Although the ELSA and RAND methods are not completely nested (e.g., ELSA includes interactions between the two covariates that define the cells), the RAND method appears more flexible overall and that is how we will interpret the results.

The theoretical and empirical findings in the literature, as summarized in sections 1 and 2, then suggest that the RAND method is more likely to be unbiased, but more prone to overfitting. One consequence is that the ELSA imputations will more likely be biased toward the mean (the "match bias" as discussed by Hirsch & Schumacher, 2004 and Bollinger & Hirsch, 2006), whereas the RAND method may impute too many outliers. Indeed, RAND introduced the tobit method for the top bracket because the nearest neighbor method imputed too many outliers in the top bracket (Michael Hurd, personal communication).

Another drawback of the RAND approach is the complexity in methodology, which requires a large number of submodels for different situations and fallback methods for special cases, and extensive checks. Implementing this method is very costly and therefore likely infeasible for other surveys, such as JSTAR, CHARLS, and LASI. Thus, if it turns out that analyses using the ELSA method give similar results as the ones using the RAND method, the ELSA method would be preferable for those surveys.

Because the ELSA imputation method is much simpler and thus more straightforward to replicate than the RAND method, we compare imputation methods by implementing the ELSA imputation method in the HRS data, and then comparing the results to RAND's imputations on the HRS data. We follow the ELSA method quite closely for imputing couple-level variables: we define the benefit-unit type variable and construct age bands. Because ELSA's age bands are defined using England's state pension age, we define our bands using 65 in the place of state-pension age. We use these two variables, along with the bracket information, to define donor pools for each household and randomly draw from this donor pool. We also follow ELSA by not imputing for nursing-home interviews. Therefore, we drop such interviews from the RAND sample as well. As ELSA does, we impute individual earnings for respondents using only gender, our slightly modified age bands, and bracket information. Unlike ELSA, we impute couple-level variables for households without a financial respondent. ELSA also asks couples whether they keep their finances separate. If so, both members report only their own income and assets, and ELSA does not impute income and assets for a nonresponding spouse in this case. The HRS does allow couples to report separately, and thus this difference has no consequences for our implementation. In the tables and the text below, we will use the term "hot-deck" to refer to our version of the ELSA conditional hot-deck method, in order to avoid confusion about whether we are using the ELSA data or the exact ELSA method, neither of which is the case.

In addition to the RAND and ELSA methods, we consider an intermediate one. This is inspired by RAND's main method for imputing amounts, the nearest neighbor method, but can also be viewed as an extension of ELSA's hot-deck method. This method is a special case of *predictive mean matching* (PMM; Little, 1988): we use all reported values, treating "no ownership" as a zero value, and estimate a linear regression model for the inverse hyperbolic sine of this value with a fairly broad set of covariates.

The imputations of all four variables use the following common covariates, inspired by the RAND set of explanatory variables: dummy variables for census divisions, whether the financial unit is a single person or a couple, and household size (top-coded at 4 people). The three couple-level variables use a set of financial-unit-related covariates, all of which refer to the financial respondent: gender, gender interacted with financial unit type, years of education, age, age squared, race/ethnicity, self-reported health, a cognition score, proxy interview status, and a dummy for age less than 65. Our one respondent-level variable (wages and salary income) uses the same set of covariates, but referring to the respondent whose income is imputed instead of the financial respondent. The two income variables additionally use whether the respondent is working for pay (referring to the financial respondent for dividend income) and whether the spouse of the respondent is working for pay. The two asset variables instead add the inverse hyperbolic sine of total couple-level income from wages and salary and the inverse hyperbolic sine of dividend income.

As with RAND's nearest neighbor imputation, the imputed value is then the *reported* value of the household with the closest *predicted* value, where the donor pool consists of the households who reported a value that is consistent with the reported information from the household that needs imputation. In addition to the value, ownership is also taken from the donor household, so both are jointly imputed. (Note that most data sets allow ownership with a zero value, so the two are not equivalent.)

Compared to ELSA's hot-deck method, this method uses a larger set of covariates, similar to the RAND set, but not equivalent, and without a principal components step. Compared to the RAND method, the main difference is that it uses only one imputation model instead of a complex system of equations for different steps and different situations. A priori, we would expect that the results with this PMM method would be in between the ones using the hot-deck method and the RAND method.

4. Results

In the tables and figures presented below, we compare the following samples: (1) "No imputations", the observations that report a continuous value or report that they do not own the income or asset; (2) "RAND HRS imputations", which is the sample of only the missing observations, with their RAND imputations; (3) "Hot-deck imputations", the same sample as (2), but imputed using our hot-deck method; (4) "PMM imputations", the same sample as (2), but imputed using our PMM method; (5) "RAND HRS total", the full sample (excluding the nursing home residents), which combines samples (1) and (2); (6) "Hot-deck total", which combines (1) and (3); (7) "PMM total", which combines (1) and (4).

4.1 Effects on marginal distributions and correlations

Table 2 shows summary characteristics of the marginal distributions of the four variables (Wages/salary, Dividend, Checking and savings, Stocks and mutual funds) for the different samples and imputation methods. This confirms the much higher fraction of ownership in the imputed data than in the reported data, as was already clear from Table 1. As a result, means, standard deviations, and medians are also higher in the imputed data than in the reported data.

This table shows some similarities in the marginal distributions of the imputations. For example, the means of the wages and salary imputations are very similar. But there are also sizable differences. However, there is no clear pattern in the differences across the different income and asset components. Because only a small fraction of observations needs to be imputed, the consequences for the distribution in the full sample are much more limited than for the imputations themselves, but not negligible.

The largest differences are observed with dividend income, where the mean and standard deviation with the PMM method are much higher than with the other methods. This is unexpected, because methodologically, the PMM method is approximately "in between" the other two methods. However, it can be explained by the higher tendency of the PMM method to impute outliers, which, as mentioned earlier, was the reason why RAND switched to imputing from a tobit model in the top bracket. As mentioned in the Introduction, the hot deck has the opposite tendency of imputing too many values in the middle of the distribution, the match bias.

We studied this for dividend income. The median reported value is \$3,000 and the 99th percentile is \$150,000. The five highest reported values are: \$1,000,000, \$560,000, \$340,000, \$300,000, and \$288,000. The hotdeck imputes these to 1, 0, 0, 0, and 1 household, respectively, whereas the PMM imputes these to 2, 3, 2, 0, and 1 household, respectively. So especially the highest outliers are imputed more often with the PMM method. When we exclude the highest 0.5% of reported values from the donor pool, the PMM mean for the total sample drops from \$2,631 to \$2,210, and excluding the top 1% from the donor pool reduces this further to \$2,082. This is largely a mechanical effect, of course, and truncating the distribution for the imputations is undesirable in principle, but it illustrates the effect of a small number of large amounts.

Table 3 presents a number of correlations: between RAND imputations and the hot-deck and PMM imputations, between variables imputed with the same method, and between the imputed variables and self-reported health. The top panel presents correlations of our two imputation methods with the RAND imputations. If all sets of imputations have the same systematic component (fitted values) and independent error draws from the same distribution, then these correlations would be the R-squares of the imputation models. In this interpretation, some of the correlations are on the low side. This is partly due to the fitted values not being the same, but also partly reflects that the R-square of the hot-deck (when viewed as a regression model with three fully interacted categorical variables) is likely on the low side. However, contrary to our expectations, the PMM imputations are not always more strongly correlated with the RAND imputations than the hot-deck imputations. The resulting correlations between the variables in the full sample are also lower than perhaps might be expected given the limited fraction of missing values. In particular, the correlations of the dividend income variables are low.

The correlations among the different income and wealth components in the same data set are sometimes lower than among the reported values and sometimes higher. In most cases (but not all), the correlations in the full sample are lower than among the reported values. There does not appear to be a systematic relation between the relative magnitudes of these correlations and the imputation method. Specifically, we would expect the RAND variables to have higher correlations, which is the case for some combinations of variables (e.g., stock wealth and dividend income) but not others, at least not when looking at the total sample (e.g., wages/salary income and stock wealth).

The correlations with self-reported health are substantially lower for the hot-deck imputations than for the RAND and PMM imputations, with the exception of dividend income, where the correlation is lowest for PMM. This largely reflects that health is used as a covariate in the RAND and PMM imputations but not in the hot-deck imputations. In the full sample, this is still visible, but the differences are small.

Table 4 shows the correlation of each of the four variables with its lagged value from wave 8 (2006), restricting the sample to those households that had a reported value in 2006 (including no ownership), so this abstracts from imputation in 2006. This table also shows the standard deviations of the changes (first differences) in the income and wealth components for the same subsample. Jointly, these results give an indication of the persistence and random variation across time of these variables.

In this table, we also show the corresponding results using the most recent version of the RAND HRS, version N (Chien et al., 2014). Between version L and version N, there have been some data updates and corrections, which explain the differences between the RAND HRS version L results we have been studying so far and the version N results for the income variables. For the wealth variables, the imputations have been changed considerably, with the incorporation of corrections from the asset reconciliation section (section U), in which households where asset values changed a lot were asked to confirm or correct the values, and the addition of adjacent-wave ownership and value to the set of regressors in the imputation models (cross-wave imputation). Hurd et al. (2015) provide an extensive discussion of the changes and present empirical results showing the effects of these two improvements. They conclude that the section U corrections had a large impact on the reduction of the first differences in wealth. The cross-wave imputations further reduced these first differences, but the magnitude of this effect was more modest.

There are some striking differences in this table, especially for the imputations themselves. The correlations for dividend income and stock wealth are much higher for the RAND version L imputations than for the other imputations, which is unexpected, especially for stock wealth, because the RAND version N data uses the cross-wave imputation that is expected to increase serial correlationscompared to the cross-wave method used in version L. However, the correlations for the whole sample are much more similar, except that here, the correlations for the RAND version N wealth variables do stand out compared to the other methods. As expected, this is not the case for the income data. Also, for dividend income, PMM has a substantially lower correlation than the other methods. The standard deviations of the first differences largely corroborate this pattern for the total sample, with some qualifications. For wages/salary income, there are again no sizable differences and for dividend income, PMM has the largest standard deviation. For checking and savings, the standard deviations of the three cross-sectional methods are similar, but the RAND version N standard deviation is smaller as expected. For stock wealth, the standard deviations of both RAND methods are similar, and lower than for PMM and especially hot deck.

To assess the extent to which differences between the distributions can be attributed to differences in method or to random variation within the same imputation method, we repeated the hot-deck imputations four more times with different random seeds. The top panel of Figure 1 shows the kernel density estimates of the imputed values for the “no value or bracket” category, for which we have the least information, although we do know ownership. Although there is some variation between the five hot-deck imputations, their distributions are quite similar. Generally, the distributions of the RAND imputations are also reasonably similar to the distributions of the hot-deck imputations, but there are noticeable differences, and the distributions of the five hot-deck imputations are more similar to each other than to the RAND imputations. The bottom panel of Figure 1 shows the corresponding results for the PMM imputations. However, because the nearest neighbor is usually a unique observation, the five PMM densities are nearly identical for each variable,[2] and this graph only highlights the difference with the RAND imputations. The distributions of the PMM imputations are more similar to those of the hot-deck imputations than to those of the RAND imputations.

4.2 Effects on regression models

To understand how different imputation methods might affect a substantive analysis of interest, we estimated a set of regression models using the different imputation approaches. The first uses only observations without missing data, which amounts to using *listwise deletion* or *complete cases analysis*, the default of every statistical or econometric software package in the absence of imputations. The second uses the RAND imputations, the third uses the hot-deck imputations, and the fourth uses the PMM imputations. For each of these we estimate eight regression equations: two regressions for each of the four variables we study here. In each case, the income or wealth variable is the outcome variable. The first regression of the two with this outcome variable is a binary logit regression for whether the household (or individual, in the case of wages and salary income) receives the income or owns the asset, that is, an ownership regression. The second regression for each outcome variable uses the natural logarithm of the amount as the dependent variable, and thus is restricted to the subsample where this amount is strictly positive. The regressions are done at the respondent level, even if the outcome variable is a couple-level variable. This allows us to estimate models with individual-level covariates without having to decide which member of a couple to include. Note that the sample sizes for the logit regressions using imputations are the same (all households or all individuals in the sample, depending on the outcome variable), whereas the numbers for the logit "no imputation" regressions are lower, because these exclude the observations where we do not know ownership. In the amount regressions, the "no imputation" sample sizes are again the lowest, but the imputation-based sample sizes differ because imputed ownership differs between the imputations.

The explanatory variables in all regressions are the same: education in years, a dummy for females, a quadratic in age, a dummy for being in a couple, a dummy for whether the spouse works (if in a couple), a cognition score (as provided in the RAND HRS, including imputed values), and self-reported health (a five point scale from 1=excellent to 5=poor; included as a linear covariate). Because the cognition score was not recorded for individuals younger than 65 and proxy interviews (where a proxy respondent, typically a family member, answers the questions instead of the target individual), we arbitrarily set their cognition score to zero and include dummies for proxy interviews and being younger than 65.

[2] The proper way to do multiple imputation for PMM would be to randomly select a donor from the k nearest neighbors, where $k > 1$, for example, $k = 10$. However, this would likely also change the finite-sample statistical properties of the first (single) imputation, so this may not be comparable to the main PMM and RAND methods studied here.

The linear regressions are unweighted and the standard errors are the conventional OLS standard errors, so they are not clustered and not robust to heteroskedasticity. Analogously, the logit regression results are the standard unweighted maximum likelihood results.

The results are in Tables 5–8. On the whole, the results using different methods are roughly similar, as one would expect with a limited fraction of missing values and a reasonable imputation method. However, there are some notable differences in the magnitudes of coefficients and occasionally in the significance levels. Differences are generally larger for the amount regressions than for the ownership regressions, reflecting the larger fraction missing for the amount variables as well as the typically larger uncertainty about the value than about ownership of an asset or income component.

For wages and salary income (Table 5), notable differences are the coefficient of education, which indicates a 0.8 percentage point higher return to education when using the RAND imputations than when using the hot-deck imputations, with PMM being in between. Also notable is the coefficient of the dummy whether the spouse works, which is 5 percentage points higher and significant at the 1% level with the RAND imputations but not significant at the 5% level with the hot-deck and PMM imputations.

In Table 6, we see many differences for dividend income. In the ownership regression, the coefficient for whether the spouse works is considerably higher with the hot-deck and PMM imputations than with the RAND imputations, with corresponding differences in significance. In the amount regression, several coefficients are noticeably different: education, age, whether in a couple, and whether the spouse works. This also has consequences for the significance level. Most notably, the coefficient of whether the spouse works is on the border of being significant at the 5% level with the hot-deck imputations but significant at the 0.1% level with the RAND and PMM imputations.

There are no noticeable differences in significance levels in Table 7 (checking and savings), but again several coefficients have different magnitudes that are meaningful. In this table, PMM is similar to RAND instead of similar to hot-deck, and with the coefficient of self-reported health in the amount regression, RAND is even in between hot-deck and PMM.

Finally, the regressions for the value of stocks and mutual funds in Table 8 also show several notable differences: the coefficients for education, age, whether the spouse works, and self-reported health are all substantially larger with the RAND imputations than with the hot-deck imputations and to a lesser extent the PMM imputations, and the coefficient of whether the spouse works is significant at the 5% level in the regression with the RAND and PMM imputations but not significant with the hot-deck imputations. Conversely, the coefficient of being in a couple is substantially larger with the hot-deck imputations than with the RAND and PMM imputations.

There are also differences between the "no imputation" columns and the RAND and hot-deck imputation columns. This is expected if nonresponse is not completely random. The imputations are intended to correct for selectivity bias. We discuss this in the next section.

4.3 Evidence for selective nonresponse

In Table 1, we have already seen evidence for selective nonresponse: respondents who do not receive the income or do not own the asset are less likely to report a missing value. Thus, means and other measures of central tendency will be downward biased with complete-cases analysis.

Tables 5–8 also occasionally show some evidence for selective nonresponse in the value reports, for example, the difference in the coefficient of couple status in the amount regression for dividend income.

Table 9 gives further evidence for selective nonresponse. In this table, the sample consists of all observations that had a *reported* value (including no ownership) in 2006 and breaks this sample down into those who had a reported value in 2008 and those who had a missing value in 2008. For these two subsamples (which differ by variable studied), we present results on the distribution of the same variable as reported in 2006. Table 4 showed moderately large serial correlations in these variables, so differences in the 2006 distributions of these variables are indicative of differences in the 2008 distributions.

We see that missing values in 2008 are associated with a much higher fraction of ownership in 2006 for wages and salary income, dividend income, and stocks and mutual funds. Partially as a result of this, the 2006 means are two to three times higher for these variables. However, for the standard deviations we see mixed results and skewnesses are consistently and substantially lower. This implies that the *shapes* of the distributions are very different as well. These effects may be predominantly due to the differences in ownership. For checking and savings, 2006 ownership rates are very similar for the 2008 missing values and 2008 nonmissing values. For this variable, the mean, standard deviation, and skewness are all higher for the 2008 missing values than for the 2008 nonmissing values.

4.4 Effectiveness of the conditional hot-deck and PMM imputations for recovering distributions with completely random and selective nonresponse

To assess more directly whether the conditional hot-deck and PMM imputations can represent distributions in the population of interest, we performed the following experiment for each of the four income and wealth variables:

(a) We selected the sample with completely reported observations. This is the population for our experiment.
(b) We randomly set 10% to "DK ownership" (i.e., completely missing).
(c) We then reimputed the missing values using the conditional hot-deck and PMM methods.

We then performed a similar experiment with selective nonresponse instead of completely random nonresponse:

(d) Start with the data set after (a) and divide by quartiles in the variable of interest (observations randomly allocated to quartiles in the case of ties).
(e) We again set about 10% to "DK ownership", but now this is done selectively, with higher probabilities of being missing in higher quartiles. We used 2%, 7%, 13%, and 18%, respectively for the lowest through the highest quartiles.
(f) We then reimputed using the conditional hot-deck and PMM methods.

The sample under (a) represents the target distribution for this experiment, and the samples constructed in steps (c) and (f) are two imputed datasets, after random and selective nonresponse, respectively. Because the data were reimputed, the sample sizes of the three samples are the same. If the imputation is successful, the distributions after re-imputation are similar to the distribution of the target sample after step (a).

Table 10 and Figure 2 show the results. Panel (a) of Table 10 shows that the hot-deck and PMM methods generally recover the marginal distributions well for the random missing values, although there are still some noticeable differences, especially for stocks and mutual funds with the hot-deck. In the selective nonresponse condition, ownership is a little lower in the imputed dataset than in the original dataset, and (partly for this reason) the means of the distributions are consistently lower than in the original dataset, with the exception of checking and savings wealth with PMM. Panel (b) shows the correlations with self-reported health, which are consistently lower in the hot-deck imputed data (though the differences are not large), where PMM is better at preserving the original correlations. Figure 2 shows that the distributions of the logarithm of the value conditional on ownership are very close, reinforcing the conclusion that the differences in income and wealth are mostly driven by small differences in ownership rates.

5. Discussion

Most surveys suffer from missing data, and many therefore provide imputations, which should be draws from the conditional distribution of the missing data given the observed data. For practical reasons and because of limited sample size, these conditional distributions should be estimated using a relatively parsimonious explicit or implicit model. We study the sensitivity of distributions and regression analysis results using the imputed data with respect to the imputation model specification, using data on aging, which are increasingly used by researchers and policy makers to address important scientific questions and policy issues. The concerted global effort to collect comparable data on the aging populations of many countries can yield many helpful cross-country analyses for informing policy about the likely effect of institutions. However, these datasets employ different imputation methods, and therefore assessing the sensitivity of empirical results to the choice of method informs us about whether any differences found in cross-country comparisons could be due to the choice of imputation method rather than differences between the populations. Earlier results on the sensitivity of earnings in the CPS to imputation method have only limited relevance for this, because these surveys use several innovations in questionnaire design that may ameliorate the effects of different imputation methods. Our assessment involves using a simple conditional hot-deck method resembling the imputation method in the English Longitudinal Study of Ageing (ELSA) applied to the 2008 wave of the U.S. Health and Retirement Study (HRS). We also implemented an intermediate method, predictive mean matching (PMM). We compare analyses using these two methods with the same analyses applied to the official RAND HRS data, which uses RAND's imputation method, based on a much more elaborate sequential system of regression equations.

Through direct observation and indirect evidence, we find that nonresponse is *selective*: The fraction of respondents (households) who receive the income or own the asset is higher among respondents who ultimately give a (partial) nonresponse answer than among those who provide an exact answer. We also see that the means of these variables as reported in 2006 are higher for missing observations in 2008 than for non-missing observations in 2008. Finally, there are some differences in the regression analyses when not using the missing observations at all versus when using imputed values. This selectivity means that using complete cases analysis will lead to biased results. Imputation can correct for this, which is the main reason (aside from loss of information) that the aging surveys provide imputations.

We find the results using the RAND imputations and those using the hot-deck or PMM imputations are fairly similar overall, but we also see some notable differences. The differences can be quite large when looking only at the imputations, but for applied researchers, it will be more interesting to know how much this affects analyses using the full sample. For the full sample, the differences resulting from the methods of imputation are much smaller, due to the limited amount of nonresponse (6 to 21% for the four variables we study). Nevertheless, the hot-deck and PMM imputed data have lower average amounts of checking and savings balances and stock wealth than the RAND imputed data but higher average amounts of dividend income. Dividend income is the rarest of these variables, with the lowest fraction of nonzero amounts reported. This makes it most sensitive to the imputation method. Detailed study affirms the match bias of the hot deck (more imputations from the middle of the distribution) and the tendency of PMM to impute outliers for this variable.

When we use these imputed variables as dependent variables in regression models, some of the coefficients are noticeably different in magnitude and occasionally different in statistical significance. This is much more pronounced in the linear regressions with the log of the value conditional on ownership as the dependent variable than in the logit regressions with a binary indicator for ownership as the dependent variable. This is partly because the fraction missing is higher for value than for ownership.

To assess the ability of the implemented hot-deck and PMM procedures to recover the target distributions, we performed an experiment in which we started with the complete data and randomly or selectively set some of the observations to missing and then re-imputed them. We find that the imputation procedures generally performed well in the random missingness arm of this experiment, but with selective nonresponse leading to a slight bias in the fraction ownership, especially for the hot deck, which leads to lower means.

Acknowledgments

We thank James Banks, Sandy Chien, Michael Hurd, Hidehiko Ichimura, Michael Moldoff, Zoë Oldfield, Albert Park, John Strauss, and David Weir for helpful discussions and clarification. This research was supported by grant number R01 AG030153 from the National Institute on Aging. The Health and Retirement Study is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

References

Abadie, A., & Imbens, G. W. (2012) A martingale representation for matching estimators. *Journal of the American Statistical Association*, *107*, 833–843.

Arokiasamy, P., Bloom, D., Lee, J., Parasuraman, S., Feeney, K., & Ozolins, M. (2012). Longitudinal Study on Aging in India: vision, design, implementation, and preliminary findings. In J. P. Smith, & M. Majmundar (Eds.), *Aging in Asia: findings from new and emerging data initiatives*. (pp. 36–74). Washington, DC: National Academy Press, .

Banks, J., Blundell, R., Levell, P., & Smith, J. P. (2015). *Life-cycle consumption patterns at older ages in the US and the UK: Can medical expenditures explain the difference?* (Working Paper No. WR-1100). Santa Monica, CA: RAND Corporation.

Banks, J., & Smith, J. P. (2012). International comparisons in health economics: evidence from aging studies. *Annual Review of Economics*, *4*, 57–81.

Benítez-Silva, H., Buchinsky, M., Chan, H. M., Rust, J., & Sheidvasser, S. (1999). An empirical analysis of the Social Security disability application, appeal, and award process. *Labour Economics*, *6*, 147–178

Bishop, J. A., Formby, J. P., & Thistle, P. D. (2003). Can earnings equations estimates improve CPS hot-deck imputations? *Journal of Labor Research*, *24*, 153–159.

Bollinger, C. R. (1998). Measurement error in the Current Population Survey: a nonparametric look. *Journal of Labor Economics*, *16*, 576–594

Bollinger, C. R., & Hirsch, B. T. (2006). Match bias from earnings imputation in the Current Population Survey: the case of imperfect matching. *Journal of Labor Economics*, *24*, 483–519.

Börsch-Supan, A., & Jürges, H. (Eds.). (2005). *The Survey of Health, Aging, and Retirement in Europe — methodology*. Mannheim, Germany: Mannheim Research Institute for the Economics of Aging (MEA).

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of Econometrics*, *5*, 3705–3843.

Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, *9*, 1–24.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge, UK: Cambridge University Press.

Chien, S., et al. (2014) *RAND HRS data documentation, Version N*. Santa Monica, CA: RAND Corporation.

Christelis, D. (2011). *Imputation of missing data in waves 1 and 2 of SHARE* (Working Paper No. 01-2011). Mannheim, Germany: SHARE Project. doi:10.2139/ssrn.1788248

David, M., Little, R. J. A., Samuhel, M. E., & Triest, R. K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association*, *81*, 29–41.

Duncan, G. J., & Hill, D. H. (1985). An investigation of the extent and consequences of measurement error in labor-economic survey data. *Journal of Labor Economics*, *3*, 508–532.

Erosa, A., Fuster, L., & Kambourov, G. (2012). Labor supply and government programs: a cross-country analysis. *Journal of Monetary Economics*, *59*, 84–107.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: the Michigan Panel Study of Income Dynamics. *Journal of Human Resources*, *33*, 251-299.

French, E. (2005). The effects of health, wealth, and wages on labor supply and retirement behavior. *Review of Economic Studies*, *72*, 395–427.

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549-576.

Gustman, A. L., & Steinmeier, T. L. (2005). The Social Security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics*, *89*, 441–463.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Hirsch, B. T., & Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics*, *22*, 689–722.

Hoynes, H. W., Hurd, M. D., & Chand, H. (1998). Household wealth of the elderly under alternative imputation procedures. In D. A. Wise (Ed.), *Inquiries in the economics of aging* (pp. 229–257). Chicago, IL: University of Chicago Press.

Hurd, M., Juster, F. T., & Smith, J. P. (2003). Enhancing the quality of data on income: recent innovations from the HRS. *Journal of Human Resources*, *38*, 758-772.

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary costs of dementia in the United States. *New England Journal of Medicine*, *368*, 1326–1334.

Hurd, M. D., Meijer, E., Moldoff, M., & Rohwedder, S. (2015). Improved wealth measures in the Health and Retirement Study: asset reconciliation and cross-wave imputation. Santa Monica, CA: RAND Corporation. (forthcoming)

Ichimura, H., Hashimoto, H., & Shimizutani, S. (2009). *Japanese Study of Aging and Retirement: JSTAR first results 2009 report* (Discussion Paper No. 09-E-047). Tokyo, Japan: Research Institute of Economy, Trade and Industry.

Juster, F. T., & Smith, J. P. (1997). Improving the quality of economic data: lessons from the HRS and AHEAD. *Journal of the American Statistical Association*, *92*, 1268–1278.

Juster, F. T., & Suzman, R. (1995). An overview of the Health and Retirement Study. *Journal of Human Resources*, *30*, S7–S56.

Kim, J. K., & Hong, M. (2012). Imputation for statistical inference with coarse data. *Canadian Journal of Statistics*, *40*, 604–618.

KLI. (2007). *User guide for 2007 KLoSA*. Sejong, South Korea: Korea Labor Institute. Retrieved July 9, 2015, from https://www.kli.re.kr/klosa/en/userguide/lst.eklosa-0200

Lillard, L., Smith, J. P., & Welch, F. (1986). What do we really know about wages? The importance of nonreporting and Census imputation. *Journal of Political Economy*, *94*, 489–506.

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, *6*, 287–296.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Marmot, M., Banks, J., Blundell, R., Lessof, C., & Nazroo, J. (Eds.). (2003). *Health, wealth and lifestyles of the older population in England: the 2002 English Longitudinal Study of Ageing.* London: Institute for Fiscal Studies.

Michaud, P.-C., Goldman, D., Lakdawalla, D., Gailey, A., & Zheng, Y. (2011). Differences in health between Americans and Western Europeans: effects on longevity and public finance. *Social Science & Medicine*, *73*, 254–263.

Moldoff, M., et al. (2014). *RAND HRS income and wealth imputation, version N*. Santa Monica, CA: RAND Corporation, Center for the Study of Aging.

NIA (2007). *Growing older in America: the Health and Retirement Study* (NIH Publication No. 07-5757). Bethesda, MD: National Institute on Aging.

Oldfield, Z. (2014). *Financial derived variables and imputation procedures*. [Part of the ELSA documentation files]. London, UK: Institute for Fiscal Studies. Retrieved July 9, 2015, from http://doc.ukdataservice.ac.uk/doc/5050/mrdoc/pdf/5050_financial_derived_variables_and_imputations_procedures.pdf

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–95.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592. (with discussion)

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*, 473–489.

Schenker, N., & Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Annals of Statistics*, *16*, 1550–1566.

St.Clair, P., et al. (2011) *RAND HRS data documentation, Version L*. Santa Monica, CA: RAND Corporation.

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.

van der Klaauw, W., & Wolpin, K. I. (2008). Social Security and the retirement and savings behavior of low-income households. *Journal of Econometrics*, *145*, 21–42.

Zhao, Y., et al. (2013). *China Health And Retirement Longitudinal Study – 2011-2012 National Baseline Users' Guide*. Beijing, China: Peking University, National School of Development. Retrieved July 9, 2015, from http://charls.ccer.edu.cn/en/page/documentation/2011_national_baseline

Table 1: Response patterns per income and asset type (%)

	Wages/salary	Dividend	Checking, savings	Stocks, mutual funds
No income or asset	66.9	76.1	14.2	71.3
Continuous value	24.5	9.0	60.8	15.3
Complete bracket: “about” value	0.5	0.6	1.1	0.4
Complete bracket: closed interval	2.5	3.1	8.3	3.6
Complete bracket: open interval	0.1	0.5	0.1	0.1
Incomplete bracket: closed interval	0.1	0.1	0.6	0.3
Incomplete bracket: open interval	0.2	0.1	0.4	0.2
Owns; no value or bracket	1.8	3.5	8.7	2.7
DK ownership	0.5	2.9	1.6	2.0
No financial respondent	0.3	0.5	0.5	0.5
Nursing home resident	2.6	3.6	3.6	3.6
N	17,217	11,897	11,897	11,897

Table 2: Effects of imputation method on marginal distributions

	N	Mean ($)	s.d. ($)	Skewness	Median ($)	Ownership (%)
Wages, salary						
No imputations	15,741	10,314	26,990	6.5	0	26.8
RAND HRS imputations	1,035	30,501	33,242	2.1	24,000	90.6
Hot-deck imputations	1,035	30,621	41,557	7.3	22,000	90.2
PMM imputations	1,035	30,743	35,530	2.5	22,000	91.0
RAND HRS total	16,776	11,560	27,843	5.9	0	30.7
Hot-deck total	16,776	11,567	28,528	6.8	0	30.7
PMM total	16,776	11,575	28,026	5.9	0	30.8
Dividend						
No imputations	10,128	1,414	15,099	38.7	0	10.6
RAND HRS imputations	1,336	7,513	23,735	14.4	1,000	75.7
Hot-deck imputations	1,336	8,743	35,371	18.1	1,000	73.6
PMM imputations	1,336	11,851	53,697	12.6	1,000	74.9
RAND HRS total	11,464	2,125	16,457	31.6	0	18.2
Hot-deck total	11,464	2,268	18,779	32.2	0	17.9
PMM total	11,464	2,631	23,418	27.5	0	18.1
Checking, savings						
No imputations	8,928	22,984	73,146	11.2	3,750	81.0
RAND HRS imputations	2,536	38,394	90,664	7.2	7,000	97.7
Hot-deck imputations	2,536	30,440	76,156	6.3	6,000	98.3
PMM imputations	2,536	32,400	84,845	7.9	6,000	97.8
RAND HRS total	11,464	26,393	77,623	9.9	5,000	84.7
Hot-deck total	11,464	24,633	73,884	10.0	4,000	84.8
PMM total	11,464	25,067	75,986	10.2	4,100	84.7
Stocks, mutual funds						
No imputations	10,300	45,817	341,596	24.7	0	17.7
RAND HRS imputations	1,164	217,867	536,869	10.7	50,000	79.9
Hot-deck imputations	1,164	195,684	534,680	9.1	40,000	80.0
PMM imputations	1,164	202,389	623,491	11.0	45,000	80.2
RAND HRS total	11,464	63,286	369,841	21.0	0	24.0
Hot-deck total	11,464	61,034	368,636	20.6	0	24.0
PMM total	11,464	61,715	382,774	20.8	0	24.0

Note. RAND HRS is version L (2011).

Table 3: Effects of imputation method on correlations

	Wages/salary	Dividend	Checking, savings	Stocks, mutual funds
Correlations between RAND and hot-deck imputed variables				
Imputations	0.5229	0.2450	0.4268	0.4891
Total	0.9465	0.7441	0.8436	0.8907
Correlations between RAND and PMM imputed variables				
Imputations	0.6100	0.2154	0.3836	0.6269
Total	0.9634	0.6225	0.8213	0.9083
Correlations within dataset				
Correlations with Dividend				
No imputations	0.0118			
RAND HRS imputations	0.0964			
Hot-deck imputations	0.0071			
PMM imputations	0.1153			
RAND HRS total	0.0000			
Hot-deck total	-0.0024			
PMM total	-0.0051			
Correlations with Checking, savings				
No imputations	0.0289	0.1997		
RAND HRS imputations	0.1908	0.0450		
Hot-deck imputations	0.0724	-0.0080		
PMM imputations	0.1649	0.0774		
RAND HRS total	0.0230	0.1509		
Hot-deck total	0.0297	0.1381		
PMM total	0.0353	0.1361		
Correlations with Stocks, mutual funds				
No imputations	0.0264	0.7417	0.2405	
RAND HRS imputations	0.2034	0.2042	0.1935	
Hot-deck imputations	0.1605	0.1792	0.0902	
PMM imputations	0.1359	0.1702	0.0811	
RAND HRS total	0.0167	0.5903	0.2169	
Hot-deck total	0.0263	0.5267	0.2069	
PMM total	0.0228	0.4760	0.1996	
Correlations with Self-reported health				
No imputations	-0.1854	-0.0490	-0.1047	-0.0618
RAND HRS imputations	-0.1440	-0.0591	-0.0887	-0.1085
Hot-deck imputations	-0.0733	-0.0166	-0.0456	-0.0213
PMM imputations	-0.1523	-0.0022	-0.0949	-0.0719
RAND HRS total	-0.1869	-0.0578	-0.1012	-0.0743
Hot-deck total	-0.1789	-0.0477	-0.0930	-0.0618
PMM total	-0.1870	-0.0411	-0.1027	-0.0684

Note. RAND HRS is version L (2011).

Table 4: Effects of imputation method on correlations with the same variable in 2006 and standard deviation of the change in this variable (when reported in 2006)

	Wages/salary	Dividend	Checking, savings	Stocks, mutual funds
Serial correlation				
No imputations	0.3109	0.4094	0.3415	0.4459
RAND HRS (ver.L) imputations	0.3861	0.3685	0.0193	0.6796
RAND HRS (ver.N) imputations	0.4001	0.1886	0.0257	0.4449
Hot-deck imputations	0.2976	0.1911	0.0193	0.3163
PMM imputations	0.4441	0.1069	0.0106	0.5766
RAND HRS (ver.L) total	0.3100	0.4025	0.1590	0.4677
RAND HRS (ver.N) total	0.3079	0.3762	0.2327	0.5330
Hot-deck total	0.3035	0.3806	0.1585	0.4336
PMM total	0.3108	0.3108	0.1552	0.4572
s.d. of first differences				
No imputations	57,336	14,495	85,760	811,194
RAND HRS (ver.L) imputations	34,939	18,161	319,275	936,331
RAND HRS (ver.N) imputations	39,635	29,978	316,495	1,174,124
Hot-deck imputations	45,950	22,443	320,591	1,225,816
PMM imputations	34,354	51,475	323,094	1,035,334
RAND HRS (ver.L) total	56,691	14,819	140,563	819,735
RAND HRS (ver.N) total	56,815	16,204	133,783	820,535
Hot-deck total	56,980	15,256	140,955	843,521
PMM total	56,680	19,871	141,704	827,229

Note. Differences between versions L (2011) and N (2014) of the RAND HRS include much more than just imputation method. See text for explanation.

Table 5: Effects of imputation method on regressions for wages/salary income.

	Ownership				Log value			
	No Impu	RAND	Hot-deck	PMM	No Impu	RAND	Hot-deck	PMM
Education (years)	0.0637***	0.0623***	0.0624***	0.0650***	0.0826***	0.0861***	0.0776***	0.0833***
	(8.82)	(8.68)	(8.71)	(9.05)	(-12.89)	(-14.78)	(-13.26)	(-14.41)
Female	-0.326***	-0.317***	-0.324***	-0.332***	-0.456***	-0.428***	-0.452***	-0.460***
	(-7.78)	(-7.61)	(-7.78)	(-7.98)	(-13.00)	(-13.13)	(-13.82)	(-14.28)
Scaled Age	-1.052***	-1.051***	-1.047***	-1.047***	-0.485***	-0.421***	-0.454***	-0.458***
	(-24.07)	(-24.20)	(-24.11)	(-24.12)	(-11.29)	(-10.86)	(-11.69)	(-11.96)
Scaled Age Squared	-0.232***	-0.233***	-0.232***	-0.236***	-0.172***	-0.154***	-0.161***	-0.164***
	(-11.15)	(-11.22)	(-11.19)	(-11.41)	(-8.81)	(-8.97)	(-9.37)	(-9.68)
Couple	-0.388***	-0.376***	-0.381***	-0.386***	0.00967	-0.00671	-0.00164	0.0349
	(-7.69)	(-7.50)	(-7.59)	(-7.70)	(-0.21)	(-0.16)	(-0.04)	(-0.83)
Spouse works	0.502***	0.500***	0.500***	0.502***	0.062	0.0994**	0.0454	0.0462
	(10.20)	(10.23)	(10.22)	(10.28)	(-1.5)	(-2.63)	(-1.19)	(-1.23)
Cognition	0.00922	0.00986*	0.00913	0.00933	0.000244	0.000212	0.000787	0.00283
	(1.91)	(2.06)	(1.91)	(1.95)	(-0.05)	(-0.06)	(-0.2)	(-0.75)
Self-reported health	-0.319***	-0.317***	-0.313***	-0.314***	-0.142***	-0.128***	-0.124***	-0.131***
	(-16.27)	(-16.26)	(-16.08)	(-16.15)	(-7.94)	(-7.81)	(-7.52)	(-8.05)
Proxy interview	0.0699	0.0722	0.0635	0.0634	0.276**	0.307***	0.318***	0.352***
	(0.59)	(0.61)	(0.54)	(0.53)	(-2.75)	(-3.49)	(-3.59)	(-4.04)
Less than 65	0.529***	0.539***	0.537***	0.533***	0.319*	0.345**	0.402***	0.416***
	(4.00)	(4.11)	(4.10)	(4.07)	(-2.55)	(-3.2)	(-3.71)	(-3.89)
Constant	-0.450**	-0.462**	-0.459**	-0.475**	9.217***	9.099***	9.178***	9.075***
	(-2.77)	(-2.86)	(-2.84)	(-2.94)	(-60.87)	(-67.78)	(-67.84)	(-68.03)
N	16,531	16,678	16,678	16,678	4,185	5,115	5,108	5,118
(Pseudo) R-Squared	0.250	0.249	0.249	0.249	0.236	0.224	0.231	0.241

Note. RAND HRS is version L (2011); t statistics in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6: Effects of imputation method on regressions for dividend income.

	Ownership				Log value			
	No Impu	RAND	Hot-deck	PMM	No Impu	RAND	Hot-deck	PMM
Education (years)	0.256***	0.250***	0.247***	0.253***	0.166***	0.142***	0.113***	0.177***
	(28.10)	(28.07)	(27.69)	(28.20)	(-7.18)	(-8.77)	(-6.93)	(-10.83)
Female	0.113*	0.103*	0.101*	0.0950*	0.0792	0.0667	0.0723	0.0975
	(2.50)	(2.33)	(2.28)	(2.15)	(-0.75)	(-0.86)	(-0.93)	(-1.26)
Scaled Age	0.480***	0.481***	0.464***	0.471***	0.306*	0.305***	0.238**	0.287**
	(9.55)	(9.75)	(9.41)	(9.47)	(-2.37)	(-3.38)	(-2.59)	(-3.1)
Scaled Age Squared	-0.00154	-0.000219	-0.000757	-0.00548	-0.0284	0.025	0.0211	0.00891
	(-0.08)	(-0.01)	(-0.04)	(-0.29)	(-0.57)	(-0.73)	(-0.6)	(-0.25)
Couple	0.687***	0.678***	0.668***	0.661***	0.537***	0.316***	0.499***	0.381***
	(12.72)	(12.77)	(12.55)	(12.46)	(-4.04)	(-3.3)	(-5.18)	(-3.98)
Spouse works	-0.134*	-0.103	-0.120*	-0.126*	-0.258	-0.335***	-0.197	-0.327**
	(-2.39)	(-1.87)	(-2.17)	(-2.27)	(-1.87)	(-3.35)	(-1.96)	(-3.25)
Cognition	0.0298***	0.0297***	0.0299***	0.0306***	0.00249	0.0143	0.00987	0.00671
	(6.20)	(6.31)	(6.35)	(6.49)	(-0.18)	(-1.6)	(-1.09)	(-0.74)
Self-reported health	-0.260***	-0.258***	-0.250***	-0.252***	-0.191***	-0.210***	-0.168***	-0.186***
	(-12.13)	(-12.29)	(-11.85)	(-11.97)	(-3.51)	(-5.48)	(-4.34)	(-4.80)
Proxy interview	0.421**	0.389**	0.402**	0.409**	0.0747	0.345	0.039	0.308
	(3.27)	(3.09)	(3.20)	(3.24)	(-0.2)	(-1.4)	(-0.16)	(-1.24)
Less than 65	0.675***	0.680***	0.676***	0.662***	-0.343	0.0651	-0.1	-0.372
	(4.62)	(4.75)	(4.71)	(4.61)	(-0.86)	(-0.24)	(-0.36)	(-1.35)
Constant	-5.465***	-5.394***	-5.378***	-5.427***	5.446***	5.547***	5.897***	5.259***
	(-29.03)	(-29.26)	(-29.10)	(-29.34)	(-10.89)	(-16.07)	(-16.95)	(-15.19)
N	16,077	16,618	16,618	16,618	1,704	3,258	3,210	3,234
(Pseudo) R-Squared	0.126	0.123	0.119	0.123	0.074	0.069	0.051	0.089

Note. RAND HRS is version L (2011); t statistics in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 7: Effects of imputation method on regressions for checking, savings accounts.

	Ownership				Log value			
	No Impu	RAND	Hot-deck	PMM	No Impu	RAND	Hot-deck	PMM
Education (years)	0.207***	0.204***	0.204***	0.207***	0.186***	0.185***	0.153***	0.186***
	(27.04)	(26.98)	(26.94)	(27.30)	(-26.58)	(-29.84)	(-24.58)	(-29.69)
Female	0.0492	0.0541	0.0511	0.0316	-0.0326	-0.00484	-0.028	-0.0414
	(0.94)	(1.05)	(0.99)	(0.61)	(-0.84)	(-0.14)	(-0.80)	(-1.18)
Scaled Age	0.475***	0.470***	0.468***	0.464***	0.472***	0.482***	0.431***	0.483***
	(10.36)	(10.38)	(10.33)	(10.24)	(-12.41)	(-14.04)	(-12.48)	(-13.99)
Scaled Age Squared	-0.0166	-0.0135	-0.0211	-0.0180	0.0323*	0.00512	0.0136	0.0197
	(-0.97)	(-0.79)	(-1.24)	(-1.06)	(-2.17)	(-0.39)	(-1.03)	(-1.49)
Couple	0.475***	0.473***	0.506***	0.468***	0.984***	0.917***	0.996***	0.911***
	(8.32)	(8.44)	(8.98)	(8.34)	(-21.59)	(-22.69)	(-24.46)	(-22.37)
Spouse works	0.365***	0.382***	0.337***	0.372***	0.132**	0.112*	0.113*	0.101*
	(4.97)	(5.27)	(4.64)	(5.13)	(-2.69)	(-2.56)	(-2.56)	(-2.27)
Cognition	0.0478***	0.0465***	0.0469***	0.0466***	0.0257***	0.0288***	0.0220***	0.0282***
	(8.37)	(8.30)	(8.34)	(8.33)	(-5.94)	(-7.59)	(-5.75)	(-7.39)
Self-reported health	-0.226***	-0.225***	-0.220***	-0.218***	-0.332***	-0.291***	-0.268***	-0.323***
	(-9.83)	(-9.92)	(-9.65)	(-9.61)	(-18.45)	(-18.13)	(-16.63)	(-19.98)
Proxy interview	0.535***	0.528***	0.525***	0.519***	0.557***	0.639***	0.502***	0.599***
	(3.97)	(3.99)	(3.96)	(3.92)	(-4.93)	(-6.58)	(-5.11)	(-6.10)
Less than 65	1.240***	1.198***	1.211***	1.212***	0.531***	0.633***	0.442***	0.632***
	(8.10)	(7.97)	(8.03)	(8.06)	(-4.23)	(-5.75)	(-3.99)	(-5.7)
Constant	-1.477***	-1.434***	-1.446***	-1.467***	5.815***	5.752***	6.187***	5.782***
	(-8.44)	(-8.32)	(-8.38)	(-8.51)	(-38.34)	(-43.27)	(-46.17)	(-43.01)
N	16,310	16,618	16,618	16,618	10,559	14,036	14,031	14,008
(Pseudo) R-Squared	0.140	0.136	0.136	0.137	0.199	0.174	0.151	0.179

Note. RAND HRS is version L (2011); t statistics in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 8: Effects of imputation method on regressions for stocks, mutual funds.

	Ownership				Log value			
	No Impu	RAND	Hot-deck	PMM	No Impu	RAND	Hot-deck	PMM
Education (years)	0.249***	0.246***	0.240***	0.246***	0.188***	0.184***	0.136***	0.175***
	(30.42)	(30.42)	(29.95)	(30.48)	(-11.62)	(-13.94)	(-10.25)	(-13.61)
Female	0.0742	0.0674	0.0731	0.0626	0.0711	0.116	0.0591	0.0817
	(1.82)	(1.68)	(1.82)	(1.56)	(-0.96)	(-1.85)	(-0.93)	(-1.34)
Scaled Age	0.422***	0.432***	0.416***	0.415***	0.382***	0.390***	0.262***	0.311***
	(9.82)	(10.07)	(9.77)	(9.71)	(-4.71)	(-5.51)	(-3.67)	(-4.55)
Scaled Age Squared	-0.00365	-0.00722	-0.00878	-0.00756	-0.00816	0.0095	0.0481	0.019
	(-0.22)	(-0.44)	(-0.54)	(-0.46)	(-0.25)	(-0.34)	(-1.73)	(-0.71)
Couple	0.722***	0.721***	0.740***	0.704***	0.360***	0.347***	0.429***	0.321***
	(14.74)	(14.89)	(15.32)	(14.56)	(-3.77)	(-4.41)	(-5.4)	(-4.21)
Spouse works	-0.0496	-0.0461	-0.0521	-0.0352	-0.185*	-0.163*	-0.14	-0.179*
	(-0.98)	(-0.93)	(-1.05)	(-0.71)	(-2.01)	(-2.06)	(-1.76)	(-2.33)
Cognition	0.0326***	0.0337***	0.0309***	0.0331***	0.00449	0.0102	0.00543	0.00942
	(7.40)	(7.76)	(7.12)	(7.62)	(-0.51)	(-1.42)	(-0.75)	(-1.34)
Self-reported health	-0.255***	-0.253***	-0.250***	-0.249***	-0.224***	-0.245***	-0.183***	-0.209***
	(-13.23)	(-13.28)	(-13.17)	(-13.12)	(-6.06)	(-7.93)	(-5.88)	(-6.95)
Proxy interview	0.366**	0.387***	0.312**	0.352**	-0.0411	0.128	-0.0825	0.184
	(3.15)	(3.39)	(2.72)	(3.07)	(-0.16)	(-0.66)	(-0.42)	(-0.97)
Less than 65	0.794***	0.828***	0.754***	0.782***	0.0795	0.0341	-0.103	-0.0196
	(6.04)	(6.38)	(5.83)	(6.02)	(-0.3)	(-0.16)	(-0.47)	(-0.09)
Constant	-5.030***	-5.028***	-4.898***	-4.996***	8.390***	8.398***	8.980***	8.512***
	(-29.68)	(-30.01)	(-29.43)	(-29.86)	(-24.47)	(-29.96)	(-31.93)	(-31.24)
N	16,240	16,618	16,618	16,618	2,921	4,298	4,301	4,296
(Pseudo) R-Squared	0.127	0.125	0.122	0.124	0.088	0.100	0.065	0.090

Note. RAND HRS is version L (2011); *t* statistics in parentheses

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 9: Marginal distributions in 2006, by response status in 2008

	N	Mean	s.d.	Skewness	Median	Ownership (%)
Wages, salary						
Reported in 2008	14,481	11,113	59,693	89.1	0	28.6
Missing in 2008	557	20,720	27,608	2.3	11,200	63.2
Dividend						
Reported in 2008	8,895	936	8,201	19.6	0	10.3
Missing in 2008	718	3,218	13,645	8.2	0	28.8
Checking, savings						
Reported in 2008	7,440	21,997	73,576	16.7	4,000	82.1
Missing in 2008	1,124	31,375	310,914	29.8	3,100	81.1
Stocks, mutual funds						
Reported in 2008	9,289	50,777	896,276	78.0	0	18.7
Missing in 2008	629	146,665	1,313,304	20.1	0	38.5

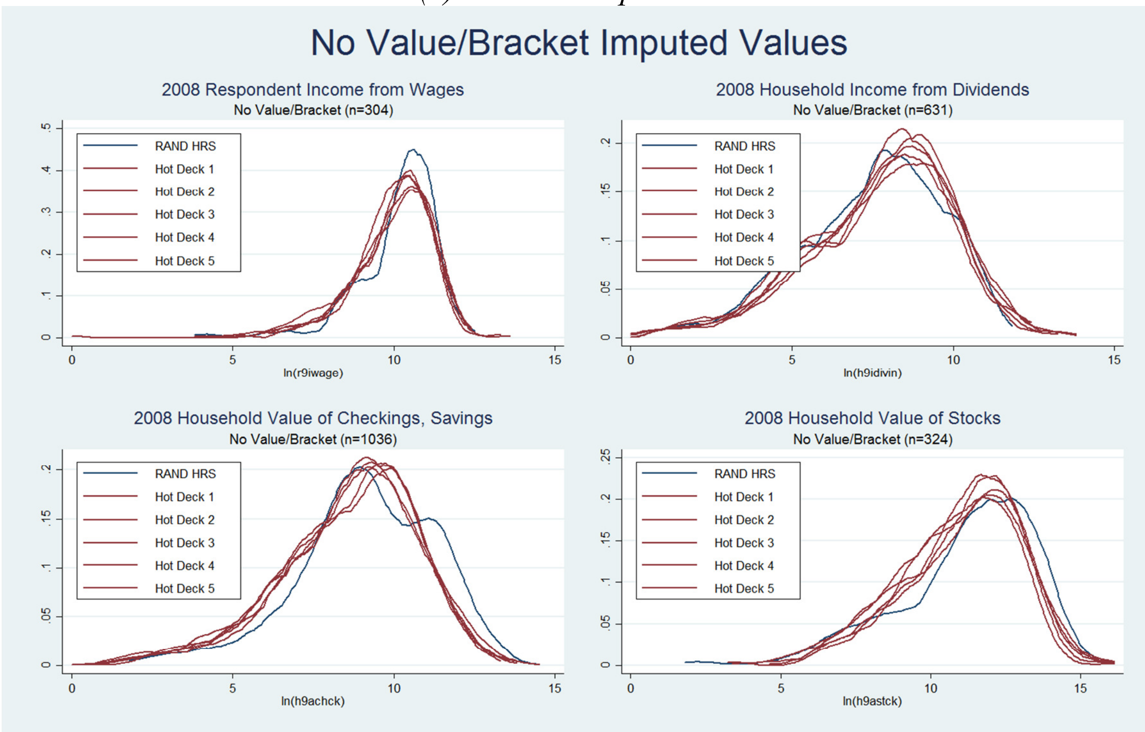
Table 10: Analyses using reported values and after setting a fraction of these to missing and reimputing with hot-deck

	(a) Marginal distributions					
	N	Mean	s.d.	Skewness	Median	Ownership (%)
Wages, salary						
Reported values	15,741	10,314	26,990	6.5	0	26.8
Reimputed:						
Hotdeck, random missings	15,741	10,327	27,326	6.9	0	26.7
PMM, random missings	15,741	10,176	26,407	6.2	0	26.8
Hotdeck, selective missings	15,741	9,542	25,867	6.6	0	25.0
PMM after selective missings	15,741	9,722	26,321	7.0	0	25.3
Dividend						
Reported values	10,128	1,414	15,099	38.7	0	10.6
Reimputed:						
Hotdeck, random missings	10,128	1,364	14,782	40.6	0	10.6
PMM, random missings	10,128	1,443	15,272	37.7	0	10.4
Hotdeck, selective missings	10,128	1,382	17,020	43.2	0	10.0
PMM after selective missings	10,128	1,394	14,977	39.3	0	10.0
Checking, savings						
Reported values	8,928	22,984	73,146	11.2	3,750	81.0
Reimputed:						
Hotdeck, random missings	8,928	22,991	71,980	10.6	3,800	80.6
PMM, random missings	8,928	23,303	74,839	11.5	4,000	81.0
Hotdeck, selective missings	8,928	20,791	67,132	11.6	3,000	79.2
PMM after selective missings	8,928	23,451	84,374	11.9	3,000	79.9
Stocks, mutual funds						
Reported values	10,300	45,817	341,596	24.7	0	17.7
Reimputed:						
Hotdeck, random missings	10,300	42,533	296,282	26.9	0	17.6
PMM, random missings	10,300	47,448	373,387	23.4	0	17.6
Hotdeck, selective missings	10,300	40,533	327,361	27.2	0	16.2
PMM after selective missings	10,300	42,555	303,549	21.6	0	17.1

	(b) Correlations of self-reported health with:			
	Wages, salary	Dividend	Checking, savings	Stocks, mutual funds
Reported values	-0.1854	-0.0490	-0.1047	-0.0618
Reimputed:				
Hotdeck, random missings	-0.1704	-0.0432	-0.0914	-0.0567
PMM, random missings	-0.1821	-0.0508	-0.1008	-0.0621
Hotdeck, selective missings	-0.1660	-0.0397	-0.0912	-0.0570
PMM after selective missings	-0.1822	-0.0454	-0.1011	-0.0688

Figure 1: Densities of imputations for “no value or bracket” observations; RAND vs. multiple hot-deck and PMM imputations.

(a) Hot-deck imputations



(b) PMM imputations

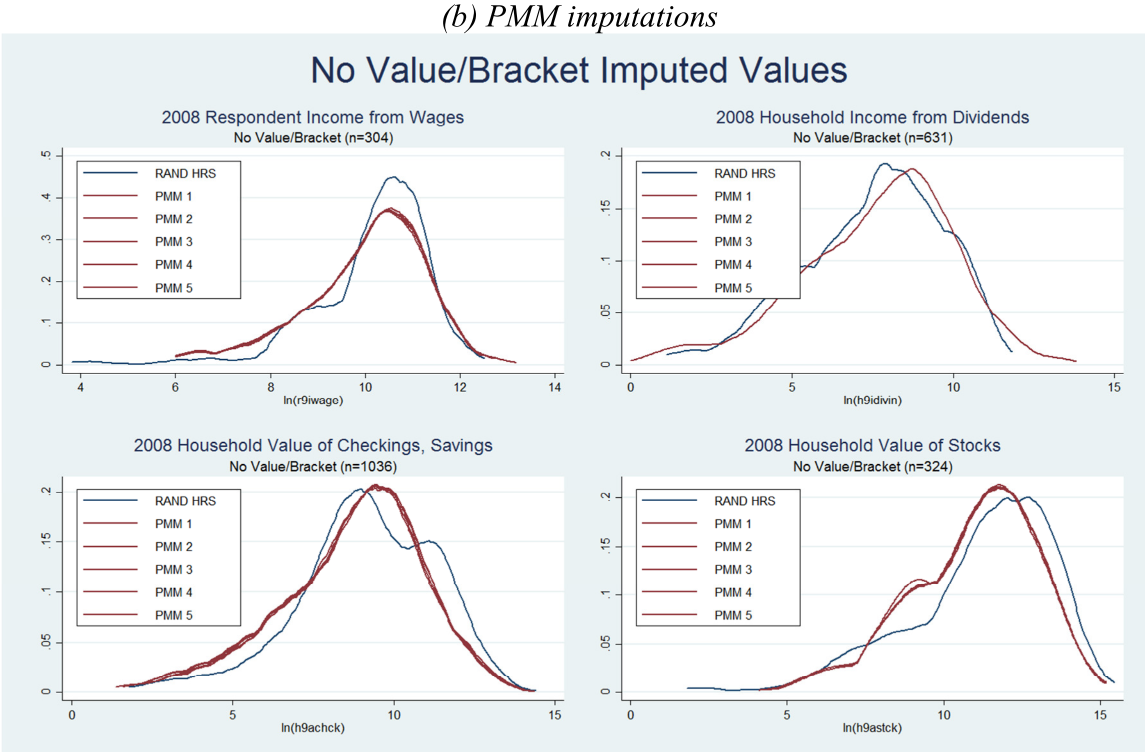
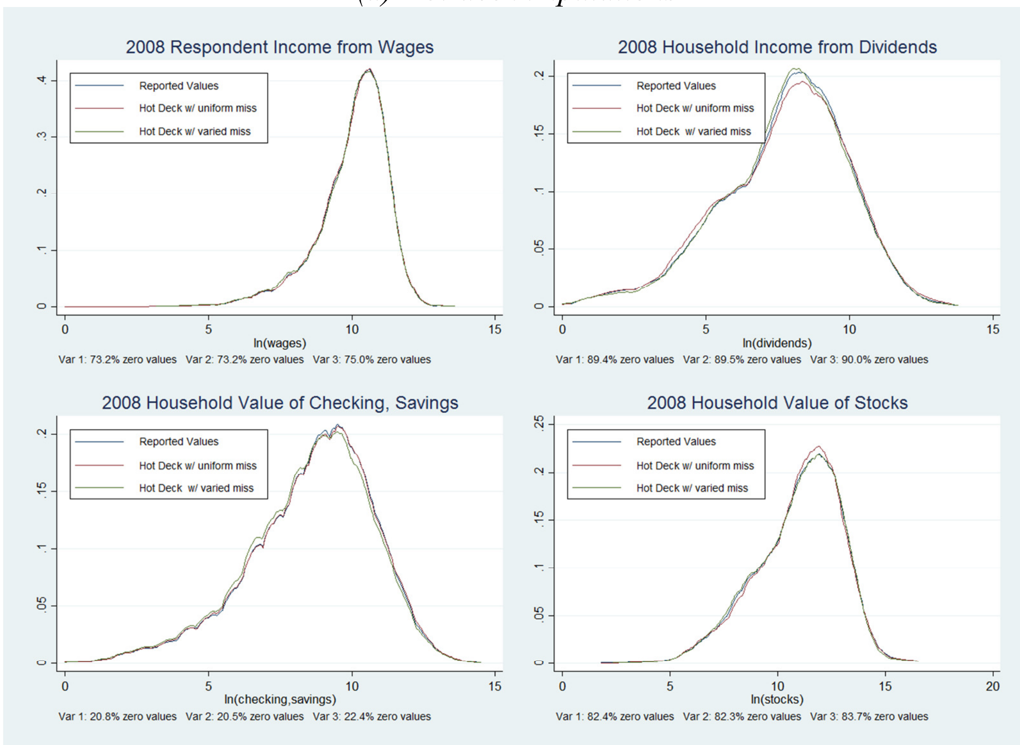


Figure 2: Ability of conditional hot-deck and PMM to recover the underlying distribution with random and selective missingness

(a) Hot-deck imputations



(b) PMM imputations