MULTIPLE IMPUTATION WITH MASSIVE DATA: AN APPLICATION TO THE PANEL STUDY OF INCOME DYNAMICS

YAJUAN SI*
STEVE HEERINGA
DAVID JOHNSON
RODERICK J. A. LITTLE
WENSHUO LIU
FABIAN PFEFFER
TRIVELLORE RAGHUNATHAN

Multiple imputation (MI) is a popular and well-established method for handling missing data in multivariate data sets, but its practicality for use in massive and complex data sets has been questioned. One such data set is the Panel Study of Income Dynamics (PSID), a longstanding and extensive survey of household income and wealth in the United States. Missing data for this survey are currently handled using traditional hot deck methods because of the simple implementation; however, the univariate hot deck results in large random wealth fluctuations. MI is

YAJUAN SI is a Research Assistant Professor, STEVE HEERINGA is a Senior Research Scientist, and DAVID JOHNSON is a Research Professor with the Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48104, USA. RODERICK J. A. LITTLE is Professor with the Department of Biostatistics, School of Public Health, 1415 Washington Heights, Ann Arbor, Michigan 48109; Research Professor, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. WENSHUO LIU is a Senior Research Scientist with the Research and Innovation, Interactions LLC, 31 Hayward Street Suite E, Franklin, MA 02038, USA. FABIAN PFEFFER is an Associate Professor with the Department of Sociology; Research Associate Professor, Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48104, USA. TRIVELLORE RAGHUNATHAN is Professor with the Department of Biostatistics, School of Public Health, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA; Research Professor, Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48104, USA

*Address correspondence to Yajuan Si, Survey Research Center, Institute for Social Research, University of Michigan, ISR 4014, 426 Thompson St., Ann Arbor, MI 48104, USA; E-mail: yajuan@umich.edu.

effective but faced with operational challenges. We use a sequential regression/chained-equation approach, using the software IVEware, to multiply impute cross-sectional wealth data in the 2013 PSID, and compare analyses of the resulting imputed data with those from the current hot deck approach. Practical difficulties, such as non-normally distributed variables, skip patterns, categorical variables with many levels, and multicollinearity, are described together with our approaches to overcoming them. We evaluate the imputation quality and validity with internal diagnostics and external benchmarking data. MI produces improvements over the existing hot deck approach by helping preserve correlation structures, such as the associations between PSID wealth components and the relationships between the household net worth and sociodemographic factors, and facilitates completed data analyses with general purposes. MI incorporates highly predictive covariates into imputation models and increases efficiency. We recommend the practical implementation of MI and expect greater gains when the fraction of missing information is large.

Statement of Significance

In this paper, we address the challenges and strategies in implement multiple imputation (MI) in a practical setting. MI is now a popular and well-established method for handling missing data; however, misunderstandings of the method exist mainly due to its practicality for use in massive data. Motivated by the Panel Study of Income Dynamics, which provides rich survey data on household income and wealth in the United States, we demonstrate our strategies handling the practical difficulties of MI in the wealth component imputation to offer detailed advice for general researchers interested in applying MI with a complex dataset.

---

1. INTRODUCTION

Multiple imputation (MI) is a useful tool for dealing with missing data, given its attractive theoretical properties, its ability to handle any pattern of missing data, and the numerous computation platforms that are available in practice. Since the initial development by Rubin (1987), MI has been successfully applied in a variety of fields for missing data and more broadly to handle related problems such as measurement error, confidentiality protection, and finite population inference (Reiter and Raghunathan 2007; Van Buuren 2012; Carpenter and Kenward 2013).

The big-data era has led to the increased availability of massive datasets. Here, we focused on massive data that have many variables, different variable types and distributions, skipped patterns, and a complex dependency structure. Loh, Eltinge, Cho, and Li (2019) argued that MI under parametric models could not be successfully implemented in these settings. Others (e.g., Little 2020) contested this assertion, and Stuart, Azur, Frangakis, and Leaf (2009) and He, Zaslavsky, Landrum, Harrington, and Catalano (2010) successfully applied chained-equation approaches to large data sets. This paper aimed to illustrate some challenges in implementing MI on a large dataset by imputing an extensive set of variables in the Panel Study of Income Dynamics (PSID). The goal of the project was to multiply impute missing data in PSID's eighteen wealth components, along with the missing values of predictors of these components, for public data release. We considered a total of 409 variables with varying amounts of missing data. We note that our application was considerably more complex than that considered by Loh et al. (2019), who restricted attention to the imputation of a single variable (amount of interest and dividend income) in the US Consumer Expenditure Survey.

The current approach implemented by PSID was a univariate hot deck imputation, where each case of item nonresponse (and bracketed response) was flagged and assigned a value that was randomly drawn from the set of reported values (or within the same bracket) with selection probabilities equal to the distribution of observed continuous values (or within the respective bracket) (for details see Pfeffer and Griffin 2017). This approach has three fundamental limitations: the current hot deck approach does not condition on covariate information, it treats each source of wealth independently, and it does not allow the user to incorporate imputation uncertainty into estimates of standard errors. Moreover, even a small percentage of the hot deck imputations results in large random wealth fluctuations between waves. Pfeffer and Griffin (2017) have found that real changes in life circumstances often account for large changes in wealth, so using MI to include additional information in the imputation will help reduce the random variation. Some form of multivariate hot deck within adjustment cells would allow for the incorporation of some associations between variables (see, e.g., Haziza and Beaumont 2007, Andridge and Little 2010), but the method is ill-suited to handling the complex multivariate pattern of missing data, the fact that different covariates may be predictive of each incomplete variable, and the need to simultaneously reflect the relationships between a large number of variables. Our goal is to address these issues by applying MI, including as predictors various socioeconomic and demographic characteristics, the values of income and other asset components, and known predictors of wealth fluctuations such as changes in household composition, employment and retirement status, health conditions, and residence.

We present our solutions to various practical difficulties that arise in this and similar big-data applications. We also evaluate the imputation quality and validity with internal diagnostics and external benchmarking data, and

demonstrate the improvements MI produces over the existing hot deck approach. The paper structure is organized as follows. Section 2 reviews the background of MI methodology and existing software for both joint and sequential regression imputations. Section 3 describes the motivating PSID study with general issues related to applications with incomplete massive datasets. We depict our imputation approaches and evaluation criteria in Section 4 and illustrate the results of the PSID application in Section 5. Section 6 concludes with discussions of the main findings and extensions.

2. BACKGROUND

MI fits models to predict the missing values with predictive covariates based on the observed values. Grounded in Bayesian methodology, MI draws model parameters from their posterior distributions and then imputes draws from a posterior predictive distribution or using predictive mean matching. The result is M completed datasets,[1] each with different draws or imputations of the missing values. Variance estimation combines the within-imputation and between-imputation variance across these M datasets, using simple MI combining rules (Rubin 1987). Although its etiology is Bayesian, MI has been shown to yield efficient estimates and inferential validity from the frequentist perspective. The proper MI model should be general enough that covers all potential analyses of interest to yield congeniality (Rubin 1996; Meng 2002).

Various MI software have been developed based on joint multivariate normal distributions, for example, PROC MI (SAS Institute Inc. 2017) in SAS, the suite of mi commands and ice in Stata (StataCorp. 2021), Amelia (King, Honaker, Joseph, and Scheve 2001; available as an independent program and an R package) and the R package norm (Schafer 1997), or a sequence of fully conditional distributions, for example, IVEware (Raghunathan, Lepkowski, van Hoewyk, and Solenberger 2001; available with multiple interfaces), the R package mice (Van Buuren and Oudshoorn 1999), and the R package mi (Gelman, Hill, Su, Yajima, Pittau, et al. 2015). MI for multilevel models is available in the R packages jomo (Quartagno, Grund, and Carpenter 2019) and pan (Schafer 2016). Flexible nonparametric Bayesian mixture models are also applied to jointly impute many incomplete categorical variables and a mixed group of categorical and continuous variables, for example, NPBayesImputeCat (Si and Reiter 2013) and MixedDataImpute (Murray and Reiter 2016). Other prediction algorithms applied to chained-equation MI include classification and regression trees (Burgette and Reiter 2010), Bayesian additive regression trees (Xu, Daniels, and Winterstein 2016), and random forests (Stekhoven and Bühlmann 2012).

---

1. We use "imputed data" to represent the imputed values of missing data, use "completed data" to represent the combination of observed data and imputed data, and use "complete data" to represent observed data.

With a large number of incomplete variables of mixed types and with various structural restrictions, developing a coherent joint imputation model is challenging. In contrast, chained-equation or sequential regression imputation approaches are both flexible and computationally attractive. This approach models the regression of each variable with missing values on all the other variables in the data set, varying the type of regression models by the type of variables being imputed. Covariates include all other variables observed or imputed for that individual. The sequence of imputing missing values overwrites previously drawn values, building interdependence among imputed values and exploiting the correlational structure among covariates. It cycles iteratively through the dataset, imputing the missing values of each variable in turn. The cycles are similar to those of a Bayesian Gibbs' sampler, but the method is only approximately Bayesian because the sequence of conditional distributions may not correspond to a coherent joint distribution (Liu, Gelman, Hill, Su, and Kropko 2014; Zhu and Raghunathan 2015).

Previous studies have applied MI to large data sets with hundreds of variables and discussed detailed steps in the implementation (e.g., Stuart et al. 2009; He et al. 2010; Azur, Stuart, Frangakis, and Leaf 2011; Drechsler 2011). We work with a similar data setting but develop a systematic process with step-by-step solutions to a broad list of problems often encountered in practice, such as data transformations, variable selection, restrictions, and diagnostics. Furthermore, we supplement the previous studies by comparing MI to hot deck imputations, examining the fraction of missing information (FMI), and conducting external evaluations.

We carry out chained-equation MI using IVEware (Raghunathan 2020), which can automatically handle issues of structural zeros, restrictions, and bounded values that are present in our application, and also has options for variable selection that are important given our large set of potential predictors. Drechsler (2011) summarizes features included in different software packages. Not all these capabilities are currently available in alternative chained-equation software. For example, the R packages mice and mi cannot directly handle skip patterns and require additional programming efforts. The IVEware imputations order variables by the amount of missing values from least to most, and draw from the posterior predictive distribution specified by the regression model with a flat or relatively noninformative prior distribution for the parameters in the regression model. Informative prior distributions can be introduced to the imputation model in mi and facilitate the variable selection. However, the PROC MI and the mi suite in Stata do not have the capabilities mentioned above.

3. MOTIVATING CASE STUDY

The PSID began in 1968 with a sample of over 18,000 individuals living in 5,000 US families, and it has followed them for the past five decades to study

the US population (PSID 2020). The PSID sample is dynamic and grows as children and grandchildren from these families form their own households and are recruited into the PSID sample and longitudinal data collection. One of the key study topics of interest to researchers is the collection of wealth information for the households of these individuals and their descendants. The survey instrument contains questions on a range of separate wealth components, such as home values, mortgages, different types of financial assets, real assets, and debts; together, these components form a measure of net worth. Item nonresponse for most wealth components is quite low in the PSID (<5 percent) partly due to the study's use of unfolding brackets and editing system to minimize nonresponse. However, about 20 percent to 25 percent of families do not report a continuous value (e.g., either report a bracketed value or no value) for at least one of the components needed to compute net worth. Consequently, since the first wealth module in 1984, the PSID has imputed missing values for users and summed the imputed components to create aggregate net worth measures.

We focused on cross-sectional PSID data for 2013, including all families who responded (9,063) and 409 variables that we selected to capture key sociodemographic characteristics of households, household reference persons, and partners, such as employment, wages, family income, consumption, race, education, and wealth. We drew on family-level and individual-level information as potential predictors of wealth components. The family-level predictors included income, consumption, and other financial measures, and the individual-level predictors included sociodemographic variables, employment, wage, and income information for household reference persons and partners (for a full list, see the Supplementary data online; PSID 2013). As part of the chained-equation regressions, all components of wealth also became predictors of other components of wealth.

In datasets with structural zeros, where certain variables are "not applicable" given values of other variables, it is important to code variables in a way that distinguishes between "not applicable" and "missing." For instance, in the case of wealth variables, PSID includes filter questions that ask whether a household holds a certain asset or not, such as whether the household owns their home. To those indicating homeownership, a follow-up question is fielded to ascertain the value of the home and the remaining mortgages. For those indicating no homeownership, the follow-up variable is recorded as "non-applicable." Both the house ownership indicator and the house value variable may be missing.

4. METHODS AND EVALUATION CRITERIA

We now describe the process of creating MI datasets based on sequential regression imputation models, from data preparation, developing imputation models, creating MIs, to model diagnostics.

4.1 Data Preparation

4.1.1 Sample design and weights. The design information (strata, clustering, and weights) may be informative for the imputation process. Methods for handling clustering variables and survey weights include models with random effects (Reiter, Raghunathan, and Kinney 2006; Quartagno, Carpenter, and Goldstein 2020), and penalized splines of propensity prediction (Zheng and Little 2003). IVEware has implemented the weighted finite population Bayesian bootstrap methods to incorporate weighting, clustering, and stratification in the synthetic population generation (Zhou, Elliott, and Raghunathan 2016). However, the computation with large-scale datasets is challenging. The sequential imputation model in IVEware provides flexibility in the selection of covariates, and we assume that the design information becomes noninformative after conditioning on the variables that are highly predictive of wealth. IVEware is yet unable to account for clustering effects and fit multilevel models. The imputation model specification and sequential implementation need novel modifications to properly account for the design information. We thus omitted the design information in the imputation process and considered it as a possible future extension in the Discussion.

4.1.2 Recoding missing values. We imputed missing values for all wealth measures and predictors. As mentioned above, different values (e.g., 999,999, 0, or NA) indicating missing data are recoded as the same missing flags. PSID also provides flags for some variables to indicate changes to the originally reported value that resulted from editing or other processing steps. Cases that are edited during data cleaning are not considered missing. If the flag code showed that the value had been imputed by other methods, we recoded it as missing. The "not applicable" values were treated as such and passed over in the imputation process via specified restrictions.

4.1.3 Outlier detection. Typically, outlier detection has been carried out in the data cleaning process to detect errors with prespecified editing rules. There are also some legitimate observations that are not errors and may be extreme and could be influential. Influential units can be handled with robust methods (e.g., Chen, Haziza, and Michal 2020). We plotted the frequency distributions of observed values to detect extreme values that would introduce skewness to the distributions of the individual variables.

4.1.4 Transformations. With legitimately extreme values, we performed transformations before imputation to limit the influence of outliers. After visually checking the frequency histograms of the observed values and calculating skewness parameters, we chose the cube-root transformation for the wealth, income, and wage variables. This transformation substantially symmetrizes the shape of the distribution. Unlike the logarithm, the cube-root transformation

can be applied to negative and zero values, which occur for some variables in our study. Congeniality between the imputation model and the analysis model is an important consideration, requiring that the imputation of inclusive of variables that could be included in an analysis model and that are associated with the outcome (Xie and Meng 2017).

4.2. Developing the Imputation Models

In fitting the imputation models, multicollinearity issues were the main source of run errors. One reason was that PSID supplemented with slightly different recoded versions of the same variables as well as versions of aggregated variables created based on user interests. Naturally, the aggregated variables are collinear with their components. Before the imputation, we applied principal component analyses to identify such variables and remove them from the set of predictors for the imputation models.

Another cause of multicollinearity was categorical variables with many nominal levels, such as the state of residence with fifty-one values. This variable was included as a predictor in the regression model with fifty-one dummy variables, leading to potential collinearity, especially from categories with small sample sizes, in this case, small states. Before the imputation process, we used forward selection to select the dummy variables along with other predictors to avoid issues with collinearity. Case identifiers, flag variables, and boundaries of intervals were not used as predictors during imputation but used in defining whether the imputation should be done or not and for the boundaries for the imputed values.

Imputations are draws from the predictive posterior distribution of each missing variable, based on a regression model for each incomplete variable, preferably with all the predictors. Given a large number of predictors, we used forward variable selection at each imputation cycle to identify a subset of the predictors tailored to the variable type: specifically, linear for continuous variables, logistic or multinomial logistic for categorical variables, and Poisson for count variables. For semicontinuous variables, such as wealth, income, and consumption, a two-stage model was used to impute missing values of a mix of a binary variable indicating presence or absence and a continuous value if the variable is present. First, a logistic regression model was used to impute the presence or absence of an asset. Conditional on imputing a nonzero status, a normal linear regression model for the cube-root transformed outcome value was then used to impute nonzero values. For example, we first imputed the indicator of whether the family had any real estate and then imputed the real estate value if owned. The indicators of nonzero status and amounts if nonzero then became potential predictors in imputation models for other variables.

Eliminating covariates that were not predictive of the outcome variable with missing values avoided problems with multicollinearity and helped to improve

the convergence of the IVEware iterations. At each imputation cycle, IVEware has the ability to use the marginal increase of the goodness of fit statistic R^2 when including each variable to select the variable. We set a minimum R^2 increase of 0.005 and a maximum number of predictors of 10. If the categorical variable is declared as a single variable, the variable with multiple categories is jointly tested, excluded, or included. If dummy variables are created and listed separately, each individual dummy variable is tested. The imputation model selects variables strongly correlated with the outcome and improves efficiency; inclusion of variables not related to the outcome will inflate variances (Little and Vartivarian 2005).

The forward variable selection method to determine the models is admittedly rather ad hoc, but perhaps justified given the size and complexity of the problem, the need to avoid collinearities, and the goal of prediction of the missing values rather than interpretation as in a substantive model. Hot deck imputation with appropriately defined adjustment cells is a possible alternative, but regression-based methods are much more flexible about conditioning on an extensive set of covariates than the hot dock. The hot deck has an implicit model that, like any model, needs to be checked (David, Little, Samuhel, and Triest 1986). In principle, more sophisticated methods such as ridge regression or lasso could be implemented (see, e.g., Deng, Chang, Ido, and Long 2016), although doing so would be challenging in this particular missing data setting. Often additive models are the default option in regression models, but interactions that are potentially predictive can and should be explicitly included as covariates. We can use tree-based methods that can detect interactions and nonlinear relationships as imputation approaches. IVEware needs extensions to allow rigorous variable selection and flexible imputation algorithms.

All regression models were fitted only to the set of applicable cases; for example, the regression model for the imputation of house values was restricted to households that own their homes. Because homeownership can itself be imputed, the set of applicable cases changes in each imputation iteration. These restrictions can be nested and must be explicitly specified so that the higher-level restricting variables are not used as predictors in the regression model. Restrictions also arise from nested skip patterns in the questionnaire. For example, a question about a loan for a second house is asked only when the respondent indicates first having a house and then a second house. Given specified restrictions, some variables are constrained to be positive, for example, the value of an owned house. However, other variables could still take on 0 values with restrictions, requiring a semicontinuous variable declaration. For example, a family could own real estate (restriction) but with or without a mortgage, where the mortgage value is a semicontinuous variable.

Some missing values come with logical or consistency bounds that must be accounted for during imputation. For example, the annual property tax or insurance premium amount must be non-negative. For some survey variables such as wealth and income, some respondents did not provide an exact value and

instead answered follow-up questions asking for brackets or range for survey variable, which then defined bounds within which the imputed values must lie. The bounds are incorporated by drawing imputations from a predictive distribution restricted to lie within the bounds, which is an option in IVEware.

4.3. Creating Imputations from IVEware Multiple Chains

The number of cycles in the initial burn-in period, the number of iterations between creating one set of imputations, and the number M of imputations to be performed, together determine the total number of cycles of the Gibbs' type algorithm.

4.4. Model Diagnostics

It is important to check that imputations are plausible as unanticipated problems in setting up the regressions can lead to poor imputations. Imputed and observed values can be compared using graphical and numeric diagnostic tools (Stuart et al. 2009). The marginal distributions of observed and imputed values are expected to be similar under missing completely at random but may markedly differ under missing at random conditions. Nevertheless, comparisons across the imputed data sets will be useful as the first phase of evaluation. Bondarenko and Raghunathan (2016) developed graphical and numeric diagnostic tools for MI to compare the distributions of imputed and observed values conditional on the response propensity score. However, such tools need the extension to work for imputation with restrictions and semicontinuous variables. A useful feature of MI is that the FMI, which estimates the relative increase in variance due to missing data (Rubin 1987; Raghunathan 2016), is readily computed as a simple function of the between-imputation and within-imputation variance. Note that no comparable measure is available from a single imputation method. A large value of FMI indicates a substantial increase in variance due to nonresponse and a high level of uncertainty about the imputation process.

The current hot deck imputation method performs univariate imputations, ignoring relationships with other variables, although other hot deck imputation approaches within adjustment cells are possible. MI with multivariate approaches, on the other hand, helps preserve the dependency structure by having the potential to include a large number of predictive covariates in the imputation model. To assess the validity of our imputations, we compared the imputed values produced by the existing hot deck approach and those produced by our MI approach in three ways: We assessed bivariate associations between wealth components as well as between net worth and other economic correlates, including household income, age, and education. We then assessed the performance of our newly imputed net worth variable as a prediction

outcome in a multivariate regression model with covariates that include total household income, education, race/ethnicity, age, and marital status. Finally, we compared the distribution of our newly imputed net worth variable to external benchmarking data.

5. RESULTS

In the 2013 PSID wave, 213 out of the 409 variables we selected were incomplete. The "apparent" missingness proportions—the share of cases without a valid value—varied between 0.01 percent and 99.72 percent, with a median value of 47.22 percent, and forty variables had more than 80 percent missing values. However, the apparent missingness summary is misleading as it fails to distinguish meaningful missingness arising from "non-applicable" questions. Table 1, therefore, shows the apparent and true missingness proportions for all eighteen wealth components. The latter values are much smaller than the former, indicating that most of the missing values arise from nonapplicable cases. Figure 1 depicts the missingness patterns of the components for the 1,457 families with incomplete wealth information, showing neither a monotone nor nested structure.

Most of the wealth variables, especially asset/debt amounts, are severely right-skewed, with a few very large values. Figure 2 shows the frequency histograms for the original home values (restricted to the sample families that report owning a house/apartment) and after cube-root transformation. The cube-root transformed values are approximately normal with a skewness of 0.7. The wealth, income, and wage variables are cube-root transformed in the imputation model.

Comparisons of results for different choices of the number of cycles suggested that about ten cycles were sufficient for most imputations, and we created $M = 10$ completed datasets. As diagnostics after imputation, table 2 provides a summary of the procedures for empirical evaluation and the corresponding findings. Through the estimates of descriptive statistics, bivariate associations, multivariate regression models, and comparison with external data, we demonstrated the capability of MI to preserve the data dependency structure and generate plausible imputations. We found that MI reduced within-imputation variances and improved estimation efficiency over the existing single hot deck imputation method.

5.1 Summary Statistics

Table 3 examines the summary information of total wealth and compares observed and completed datasets. The summary statistics of the completed values are generally larger than those of the observed, but not dramatically so. Table 4 presents the FMI values of filter and amount variables for the 18 PSID

Table 1. Apparent Missingness Proportions (Apparent Miss) and True Missingness Proportions (True Miss) for the Eighteen Wealth Components in the 2013 Panel Study of Income Dynamics Study

	Label	Apparent miss (%)	True miss (%)
W39B4 rel loans	Amount of loans from relatives	98.68	0.39
W39B3 legal	Amount of legal bills	99.39	0.40
W39B7 debt	Amount of other debt	98.82	0.40
A24_2 mortgage2	Remaining principal mortgage 2	95.90	0.54
W2B own estate	Amount of owned on other real estates	89.25	0.57
W11B own business	Owned amount of farm or business	92.08	0.67
W2A estate	Worth of other real estates	89.48	0.81
W39B2 med	Amount of medical bills	89.83	0.81
W39B1 stu loans	Amount of student loans	75.47	1.20
W39A credit	Amount of credit or store card debt	66.60	1.31
W11A business	Worth of farm or business	92.79	1.39
A20 house	House value	50.40	1.48
W16 stock	Profit if sold non-IRA stock	89.44	2.01
A24_1 mortgage1	Remaining principal mortgage 1	66.66	2.28
W34 bond	Profit if sold bonds/insurance	91.11	2.48
W22 annuity	Value of individual retirement accounts (IRA)/annuity	81.82	2.52
W6 vehicles	Profit if sold vehicles	4.63	4.63
W28 account	Amount of all accounts	37.77	5.10

wealth components. Most of them are very small, primarily because of the low underlying missingness proportions. The last three columns in table 4 compare the percentages between the observed and imputed values of positive indicators of whether the PSID household owns a wealth component and lists the number of cases with missing indicators. The comparison does not raise red flags for most components, except for the indicator A23 (Do you have a mortgage or loan on this property?) for A24_1 (Remaining principal of the first mortgage). In the imputation model for A23, the selected predictors include the house type (such as a one-family house, a two-family house, an apartment, a mobile home, or others), total family income, wage of the household head, indicator of credit card debts, and amount of the saving account. We did not have substantive concerns about the plausibility of the imputed values, given that only 23 values were missing.
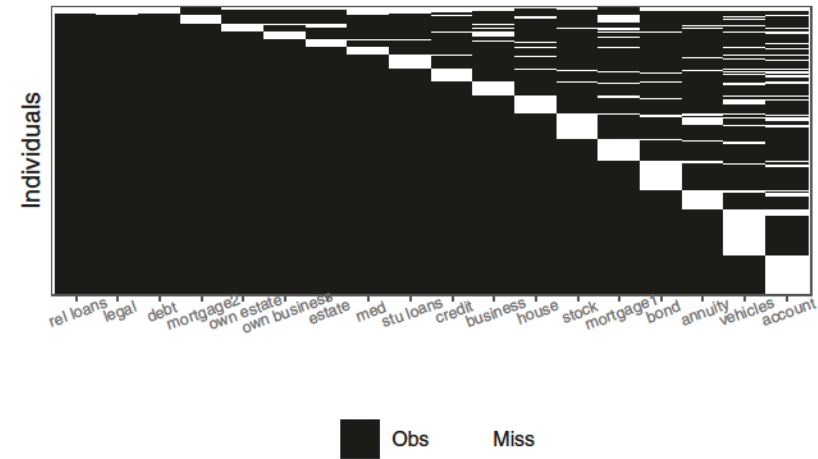
Figure 1. Observed (Obs) and Missing (Miss) Patterns of Eighteen Wealth Components of the Cases with Missing Wealth Information in the 2013 Panel Study of Income Dynamics Study.
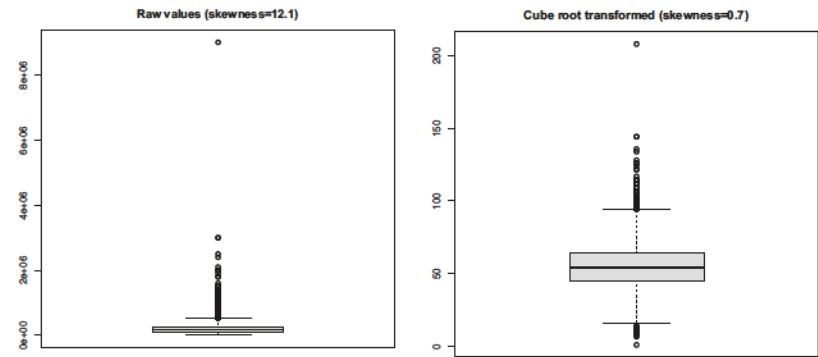

Figure 2. The Distribution and Skewness of Home Values before and after the Cube-Root Transformation.

5.2 Bivariate Associations

We first examined this MI improvement by considering the bivariate associations between wealth components. The existing imputation procedure generated the home equity value differently from MI (defined as A20 minus A24_1 minus A24_2) without reporting the three related components. This is another advantage of MI—the ability to handle hierarchical restrictions that could also be missing. Hence, we used the final home equity values, rather than the three components, and compared the resulting sixteen wealth components between the two imputation methods. Figure 3 is a scatterplot of the pairwise Pearson correlation coefficients between the sixteen wealth components, based on one

Table 2. Summary of the Empirical Evaluation Procedures and Findings in Comparison of Multiple Imputation (MI) and Hot Deck (HD) Imputation

	Findings
Summary statistics comparison between observed and imputed datasets	The summary statistics of the completed (observed and imputed data) values of MI are generally larger than those of the observed, but not dramatically so.
Fraction of missing information (FMI)	Most FMI values of filter and amount variables for the 18 PSID wealth components are very small.
Bivariate associations	MI generates larger Pearson correlation coefficient values than HD.
Multivariate regression	The coefficient estimates under different methods are similar. MI has smaller within-imputation variances and slightly larger overall variances than those under HD.
Comparison to the 2013 Survey of Consumer Finances (SCF)	The mean values from the imputation methods are similar but both are lower than the SCF estimates. The weighted estimates of business assets and other debts after MI are larger but the weighted estimates for other assets are lower than those produced by HD.

MI dataset and the dataset imputed via the hot deck. When the correlations in the observed data are not close to 0, MI generates larger values than the hot deck method, suggesting that it is preserving associations better.

Next, we evaluated the association of the household net worth with household income, age, and education. The Bland-Altman plot (Altman and Bland 1983) in figure 4 marks the data grouped by income quartiles and compares the imputed wealth values from the hot deck and one randomly selected MI dataset. The relationship between income and wealth is higher after MI than after hot deck imputation. As this relationship is expected to be positive (Killewald, Pfeffer, and Schachner 2017), this comparison favors the MI method.

Table 5 examines the ten imputed datasets and takes the average of the Pearson correlation coefficients between total net worth and total household income, age, and education, respectively. We used the cube-root transformed values for both wealth and income to dampen the effects of large outlying values. MI preserves the dependency structure from the observed data while the univariate hot deck imputation attenuates the correlation estimates. We also compared the regression coefficient estimates of these sociodemographic predictors in the univariate regression models with net worth as the outcome and

Table 3. Summary Statistics Comparison between Observed and Completed (the Combination of Observed and Imputed Data) 2013 PSID Family Wealth Values Based on One Randomly Selected Multiple Imputation Dataset: Sample Size n, Minimum/Maximum, Mean, Standard Deviation (Std), and Quantiles

	Observed	Completed	Rel. diff
n	7,606	9,063	0.19
Min	−995	−995	0
Max	33,740	33,740	0
Mean	180	200	0.11
Std	793	837	0.06
25th	0	0	
50th	14	20	0.39
75th	114	135	0.18
90th	423	487	0.15
95th	846	912	0.08

NOTE.—The relative difference (Rel.diff) is defined as (Com-Obs)/Obs. The wealth values are presented in $1,000s and rounded to the nearest 1,000.

found that the relationships in MI were larger and closer to the observed structure than those after the hot deck imputation.

Table 5 does not report the corresponding variance estimates on either the observed or imputed dataset. Because the missingness proportion is small, inferences based on the completed datasets are similar between the two imputation methods. We used table 5 to exemplify the different imputation methods' properties for preserving relationships among the imputed variables. We checked the relationship between total household income with the eighteen wealth components, respectively, and MI yields correlation coefficient estimates closer to those based on the observed values than the current hot deck method.

5.3 Multivariate Regression

We fit a multivariate regression model with cube-root transformed values of household net worth as the outcome, and covariates that included total household income (cube-root transformed), education, race/ethnicity, age, and marital status, which are some of the most commonly studied correlates of a household's wealth position (Killewald et al. 2017). Table 6 shows the coefficient estimates based on the observed values as complete case analysis, and the completed datasets after MI and hot deck imputation. We applied combining rules to the ten MI datasets to propagate the missing data uncertainty. The results under the three methods are similar, likely because the missingness proportions are small. We calculated the ratio of the MI variance estimates to

Table 4. Fraction of Missing Information for Wealth Components and the Corresponding Indicators, the Proportions (Prop) of Positive Indicators in the Observed (obs) and Imputed (imp) Data Based on One Randomly Selected Imputation, and the Number of Cases with Missing Indicators

Component	FMI	FMI of indicators	Prop. of positive indicators (Observed)	Prop. of positive indicators (Imputed)	# missing indicators
Account	0.002	0.01	0.671	0.615	65
Credit	0.006	0.003	0.344	0.318	44
Rel loans	0	0.004	0.014	0.061	33
Legal	0.011	0.003	0.006	0	33
Med	0.008	0.005	0.107	0.182	33
Estate	0.007	0.002	0.111	0.19	21
Own estate	0.012	0.002	0.111	0.19	21
Debt	0.002	0.002	0.012	0.03	33
Student loans	0.014	0.004	0.255	0.176	34
Annuity	0.01	0.002	0.203	0.2	50
Business	0.031	0.001	0.084	0.059	17
Own business	0.024	0.001	0.084	0.059	17
Bond	0.013	0.008	0.106	0.145	76
Stock	0.012	0.003	0.12	0.161	56
Vehicles	0.013	NA	NA	NA	NA
House	0.002	0	0	0	0
Mortgage1	0.023	0.002	0.355	0.739	23
Mortgage2	0.031	0.011	0.043	0.143	35

those from hot deck imputation, the overall variance, and the average within-imputation variances. MI includes the additional variance component accounting for missing data uncertainty, and the overall variances are slightly larger than those under hot deck imputation; but MI has smaller within-imputation variances, suggesting efficiency gains within imputation.

5.4 Comparison with External Data

In addition to the internal checks of imputation plausibility just presented, we compared our imputation results to external data. We contrasted our estimates with another study, the Survey of Consumer Finances (Federal Reserve Board 2020). The Survey of Consumer Finances (SCF) is a triennial cross-sectional survey of US families and collects information on families' balance sheets, pensions, income, and demographic characteristics. The SCF oversamples wealthy households, so we applied survey weights to the imputed datasets and generated population-representative estimates. Table 7 compares the mean values of key wealth components based on the 2013 PSID MI and hot deck
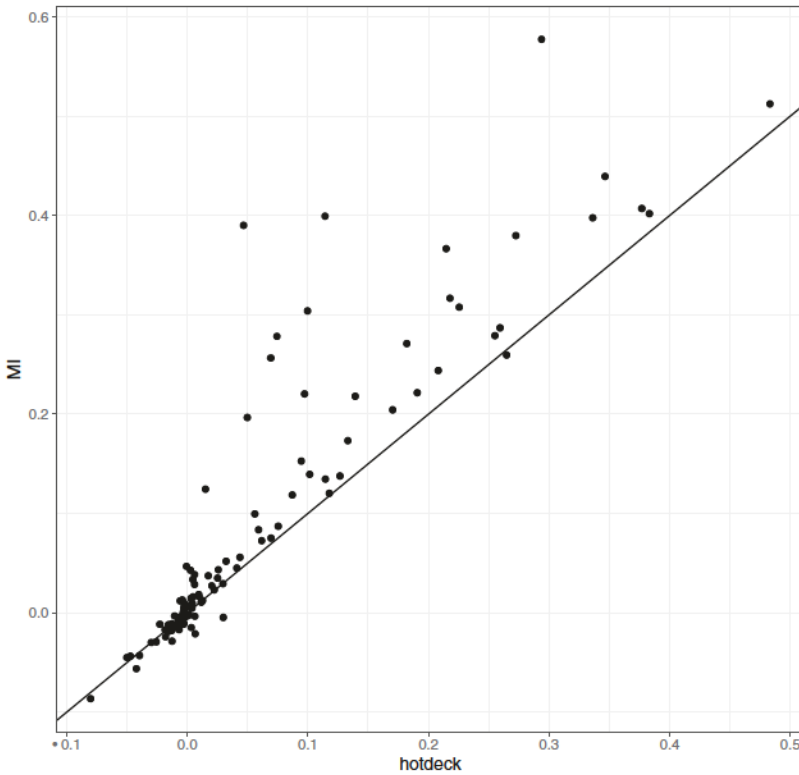
Figure 3. Pairwise Correlation Coefficients between Sixteen Wealth Components.

imputed datasets after weighting and the weighted 2013 SCF data (Bricker, Goodman, Moore, and Volz 2020; Kennickell 2020) . Because the PSID over-samples low-income families, the weighted PSID wealth estimates are higher than the unweighted. The values from the imputation methods are similar but both are lower than the SCF estimates as the SCF oversamples high-income families (Pfeffer, Schoeni, Kennickell, and Andreski 2016).

Since the wealth components have right-skewed distributions, we compare the percentiles for key distribution points: the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles, shown in Supplementary Table S1. We omitted the calculated percentiles of 0 values across all methods. Generally, the PSID estimates are lower than the SCF estimates, the latter of which have to adjust for the oversampling of wealthy families. MI generates lower values for checking/saving balances and stocks but larger values for individual retirement accounts/private annuities and other debts than the hot deck imputation. The weighted estimates of business assets and other debts after MI are larger but the weighted estimates for other assets are lower than those produced by hot deck imputation.
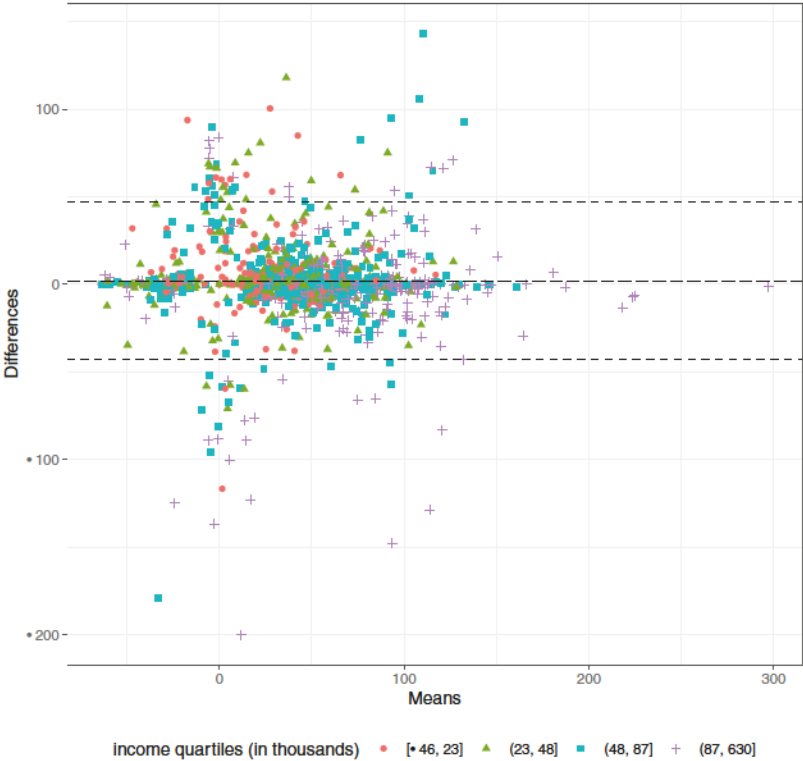
Figure 4. The Bland Altman Plot in the Comparison of Imputed Wealth Values (Cube-Root Transformed) from Multiple Imputation and Hot Deck Imputations. The three dashed lines represent the mean of differences minus two standard deviations, mean of differences and mean of differences plus two standard deviations. The critical difference (i.e., two times the standard deviation of differences) is fifty. Some wealth values are negative because of high debts and low assets.

Table 5. Correlation between Total Wealth (Cube-Root Transformed) and Sociodemographics Models in Comparison of Observed Data, Multiple Imputation and Hot Deck Imputed Data

	Observed	Multiple Imputation-Imputed	Hot Deck-Imputed
Total family income (cube-root transformed)	0.47	0.42	0.36
Age (in years)	0.41	0.37	0.35
Education (in years)	0.19	0.14	0.12

Table 6. Coefficient Estimates and Variances (Var) for the Wealth Regression Models in Comparison of Observed (Obs) Data, Completed Data (Com) from Multiple Imputation (MI) and Hot Deck (HD) Imputations

	Observed (n = 7,606)	Multiple Imputation-Completed (n = 9,063, M = 10)	Hot Deck-Completed (n = 9,063, M = 10)	Overall variance (MI/HD)	Within variance (MI/HD)
Intercept	−55.25 (−60.35, −50.15)	−55.33 (−60.14, −50.51)	−54.13 (−58.92, −49.34)	1.01	0.99
Family income	1.07 (1.01, 1.13)	1.1 (1.04, 1.16)	1.07 (1.01, 1.13)	1	1
Education (years)	0.53 (0.22, 0.85)	0.56 (0.26, 0.85)	0.57 (0.27, 0.87)	1	1
Non-Hisp black	−10.68 (−12.35, −9.01)	−11.57 (−13.12, −10.01)	−11.01 (−12.56, −9.47)	1.01	0.99
Other race/ethnicity	−7.66 (−10.1, −5.23)	−7.47 (−9.77, −5.18)	−6.74 (−9.01, −4.47)	1.03	0.98
Age	0.95 (0.89, 1.01)	0.96 (0.9, 1.02)	0.96 (0.89, 1.02)	1	1
Never married	−0.01 (−2.07, 2.05)	0.18 (−1.8, 2.16)	0.01 (−1.93, 1.95)	1.04	0.99
Widowed	−8.55 (−12.25, −4.85)	−8.17 (−11.48, −4.85)	−8.1 (−11.41, −4.79)	1	0.99
Divorced	−10.66 (−12.95, −8.37)	−10.63 (−12.81, −8.44)	−10.82 (−12.97, −8.68)	1.04	0.99
Separated	−7.58 (−11.03, −4.13)	−7.55 (−10.9, −4.2)	−7.65 (−10.94, −4.35)	1.03	0.99

NOTE.—The wealth and income values are cube-root transformed. The reference levels are non-Hispanic white for race/ethnicity and married for the marital status. The sample size is *n*, and the number of completed datasets is *M*.

Table 7. Mean Values of Total Wealth (Including Home Equities) and Key Wealth Components across Different Methods and Sources

	HD	MI	HD–wt	MI–wt	SCF
Total net worth	202,058	200,383	316,361	314,129	470,491
Business assets	29,264	29,564	48,372	46,963	110,128
Checking/savings	19,361	18,404	29,637	28,725	47,537
Stocks	32,671	32,845	56,759	57,438	67,276
IRA/private annuities	33,927	33,636	52,452	52,144	60,023
Net worth of vehicles	13,444	13,405	14,301	14,362	15,002
Equity in primary residence	58,408	58,480	86,427	86,310	103,932
Equity in real estate	18,610	18,461	28,561	28,547	52,984
Other assets	8,582	7,903	10,335	10,033	24,840
Other debts	12,209	12,316	10,484	10,393	11,230

NOTE.— MI, multiple imputation; HD, hot deck; wt, weighted; SCF, the Survey of Consumer Finances.

6. CONCLUSION AND DISCUSSION

Chained-equation imputation methods are flexible and can handle general missing data patterns of missing data. Imputations can be tailored to multiple variable types. We discussed some of the methodological and practical issues encountered when we applied chained-equation MI in a large-scale, complex, particular application, the imputation of wealth data in the PSID. We demonstrated the capability of MI to preserve the data dependency structure and improve estimation efficiency. We recommend the practical implementation of MI and call for existing MI software to be extended to accommodate the methodological issues and computational hurdles with massive datasets.

Issues addressed include (a) distinguishing missing and nonapplicable values, and tailoring the underlying regression models to applicable cases, (b) transformations of amount variables to reduce skewness, (c) simultaneous imputation of presence and amount of semi-continuous variables, (d) minimizing collinearity in the regressions, and (e) model checking. The imputation models are tailored to variable types and the imputations account for restrictions and boundaries. The forward selection was applied to restrict the predictors to variables associated with the variable being imputed, speeding the convergence of the algorithm. Multiple sequences of conditional imputation models yield $M > 1$ completed datasets that can be used for standard analyses and MI inferences that propagate imputation uncertainty.

Little (2020) summarizes the important factors in this imputation setting, such as the sample size, FMI, missingness mechanism, form and strength of the true relationship between the variable with missing data and predictors, form and strength of the true relationship between missingness and its

predictors, degree of association between the propensity to respond and the variable with missing values, and degree of misspecification of the true models for missingness and the survey design variables. We have described how we addressed each of these concerns in our imputations of PSID wealth data.

In this application, we compared MI with the existing single hot deck imputation method. To evaluate imputation plausibility, we examined descriptive summary statistics, bivariate association studies, multivariate regression models, and compared external estimates from the SCF study.

MI incorporates highly predictive covariates into imputation models and outperforms the hot deck imputations with imputation accuracy and efficiency gains. MI preserves the associations between PSID wealth components and the relationships between the household net worth and sociodemographic factors, which is crucial to such analyses with completed data. The imputation model should be more general than the analysis model, and MI facilitates completed data analyses with general purposes. As an important infrastructure survey to study intra-/intergenerational wealth and income dynamics, the PSID study will implement MI to preserve the important data correlation structure and release multiply imputed data for public use. This will substantially reduce the random variation and facilitate wealth fluctuation studies.

The detailed solutions to practical difficulties of applying MI to the PSID study also apply to large-scale studies in general. The key properties of MI are building proper imputation models that capture dependency structure of the data and propagate the imputation uncertainty. Hence, our case study offers theoretical and practical guidance to researchers implementing MI in their own data.

This comprehensive investigation invites several extensions. First, we omitted the design information in the imputation process, and further effects are necessary on practical software development to account for design information into MI with large-scale datasets. Second, we focus on the 2013 cross-sectional PSID data. The PSID's panel structure with repeated measures of the same family across time may further improve the imputation model with highly predictive variables, such as measures from previous waves of the longitudinal study, where a twofold MI procedure could be potentially useful (Welch, Petersen, Bartlett, White, Marston, et al. 2014). Third, although IVEware has a few capabilities that are not available in other MI software, further software developments are needed to fit multilevel models to account for clustering effects or correlations of longitudinal measures in imputation, specify informative prior distributions of regression coefficients to allow sophisticated variable selection, and implement flexible models or machine learning algorithms to improve the imputation performances.

Supplementary Materials

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

Altman, D. G., and J. M. Bland (1983), "Measurement in Medicine: The Analysis of Method Comparison Studies," *The Statistician*, 32, 307–317.

Andridge, R. R., and R. J. A. Little (2010), "A Review of Hot Deck Imputation for Survey Non-Response," *International Statistical Review*, 78, 40–64.

Azur, M., E. A. Stuart, C. M. Frangakis, and P. J. Leaf (2011), "Multiple Imputation by Chained Equations: What Is It and How Does It Work?," *International Journal of Methods in Psychiatric Research*, 20, 40–49.

Bondarenko, I., and T. E. Raghunathan (2016), "Graphical and Numerical Diagnostic Tools to Assess Suitability of Multiple Imputations and Imputation Models," *Statistics in Medicine*, 35, 3007–3020.

Bricker, J., S. Goodman, K. B. Moore, and A. H. Volz (2020), "Wealth and Income Concentration in the SCF: 1989–2019", available at https://www.federalreserve.gov/econres/notes/feds-notes/wealth-and-income-concentration-in-the-scf-20200928.htm. Accessed on September 8, 2021.

Burgette, L. F., and J. P. Reiter (2010), "Multiple Imputation via Sequential Regression Trees," *American Journal of Epidemiology*, 172, 1070–1076.

Carpenter, J., and M. G. Kenward (2013), *Multiple Imputation and Its Application*, Chichester: John Wiley & Sons.

Chen, S., D. Haziza, and V. Michal (2020), "Efficient Multiply Robust Imputation in the Presence of Influential Units in Surveys," available at https://arxiv.org/abs/2010.01706. Accessed on September 8, 2021.

David, M., R. J. A. Little, M. E. Samuhel, and R. K. Triest (1986), "Methods for CPS Income Imputation," *Journal of the American Statistical Association*, 81, 29–41.

Deng, Y., C. Chang, M. Ido, and Q. Long (2016), "Multiple Imputation for General Missing Data Patterns in the Presence of High-Dimensional Data," *Scientific Reports*, 6, 21689.

Drechsler, J. (2011), "Multiple Imputation in Practice: A Case Study Using a German Establishment Survey," *Advances in Statistical Analysis*, 96, 1–26.

Federal Reserve Board (2020), *Survey of Consumer Finances*, Chicago, IL: National Opinion Research Center, University of Chicago, available at https://www.federalreserve.gov/econres/scfindex.htm. Accessed on September 8, 2021.

Gelman, A., J. Hill, Y.-S. Su, M. Yajima, M. Pittau, B. Goodrich, Y. Si, and J. Kropko (2015), Package mi: Missing Data Imputation and Model Checking, R CRAN, available at https://cran.r-project.org/web/packages/mi. Accessed on September 8, 2021.

Haziza, D., and J. Beaumont (2007), "On the Construction of Imputation Classes in Surveys," *International Statistical Review*, 75, 25–43.

He, Y., A. M. Zaslavsky, M. B. Landrum, D. P. Harrington, and P. Catalano (2010), "Multiple Imputation in a Large-Scale Complex Survey: A Guide," *Statistical Methods in Medical Research*, 19, 653–670.

Kennickell, A. B. (2020), "Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research," available at https://www.federalreserve.gov/econresdata/scf/files/measurement.pdf. Accessed on September 8, 2021.

Killewald, A., F. T. Pfeffer, and J. Schachner (2017), "Wealth Inequality and Accumulation," *Annual Review of Sociology*, 43, 379–404.

King, G., J. Honaker, A. Joseph, and K. Scheve (2001), "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *American Political Science Review*, 95, 49–69.

Little, R. J. A. (2020), "On Algorithmic and Modeling Approaches to Imputation in Large Data Sets," *Statistica Sinica*, 30, 1685–1696.

Little, R. J. A., and S. Vartivarian (2005), "Does Weighting for Nonresponse Increase the Variance of Survey Means?," *Survey Methodology*, 31, 161–168.

Liu, J., A. Gelman, J. Hill, Y. Su, and J. Kropko (2014), "On the Stationary Distribution of Iterative Imputations," *Biometrika*, 101, 155–173.

Loh, W.-Y., J. Eltinge, M. J. Cho, and Y. Li (2019), "Classification and Regression Trees and Forests for Incomplete Data from Sample Surveys," *Statistica Sinica*, 29, 431–453.

Meng, X. L. (2002), "A Congenial Overview and Investigation of Multiple Imputation Inferences under Uncongeniality," in *Survey Nonresponse*, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. Little, pp. 343–356, New York, NY: John Wiley & Sons.

Murray, J. S., and J. P. Reiter (2016), "Multiple Imputation of Missing Categorical and Continuous Values via Bayesian Mixture Models with Local Dependence," *Journal of the American Statistical Association*, 111, 1466–1479.

Pfeffer, F. T., and J. Griffin (2017), "Determinants of Wealth Fluctuation: Changes in Hard-To-Measure Economic Variables in a Panel Study," *Methods, Data, Analyses*, 11, 87–108.

Pfeffer, F. T., R. Schoeni, A. Kennickell, and P. Andreski (2016), "Measuring Wealth and Wealth Inequality: Comparing Two U.S. Surveys," *Journal of Economic and Social Measurement*, 41, 103–120.

PSID (2013), *The Panel Study of Income Dynamics Questionnaire of the 2013 Data Collection, Produced and Distributed by the Survey Research Center*, Ann Arbor, MI: Institute for Social Research, University of Michigan, available at ftp://ftp.isr.umich.edu/pub/src/psid/question-naires/q2013.pdf. Accessed on September 8, 2021.

PSID (2020), *The Panel Study of Income Dynamics, Restricted Use Data, Produced and Distributed by the Survey Research Center*, Ann Arbor, MI: Institute for Social Research, University of Michigan, available at https://psidonline.isr.umich.edu. Accessed on September 8, 2021.

Quartagno, M., M. Carpenter, and J. R. Goldstein (2020), "Multiple Imputation with Survey Weights: A Multilevel Approach," *Journal of Survey Statistics and Methodology*, 8, 965–989.

Quartagno, M., S. Grund, and J. R. Carpenter (2019), "Jomo: A Flexible Package for Two-Level Joint Modelling Multiple Imputation," *The R Journal*, 11, 205–228.

Raghunathan, T. E. (2016), *Missing Data Analysis in Practice*, Boca Raton, FL: CRC Press.

———. (2020), "IVEware: Imputation and Variance Estimation Software, Version 0.3", available at https://www.src.isr.umich.edu/software/. Accessed on September 8, 2021.

Raghunathan, T. E., J. M. Lepkowski, J. van Hoewyk, and P. Solenberger (2001), "A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models," *Survey Methodology*, 27, 85–96.

Reiter, J. P., and T. E. Raghunathan (2007), "The Multiple Adaptations of Multiple Imputation," *Journal of the American Statistical Association*, 102, 1462–1471.

Reiter, J. P., T. E. Raghunathan, and S. K. Kinney (2006), "The Importance of Modeling the Survey Design in Multiple Imputation for Missing Data," *Survey Methodology*, 32, 143–150.

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York, NY: John Wiley & Sons.

———. (1996), "Multiple Imputation after 18+ Years," *Journal of the American Statistical Association*, 91, 473–489.

SAS Institute Inc. (2017), *SAS/STAT 9.4 User's Guide*, Cary, NC: SAS Institute Inc.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall.

———. (2016), Package Pan: Multiple Imputation for Multivariate Panel or Clustered Data, R CRAN, available at https://cran.r-project.org/web/packages/pan/. Accessed on September 8, 2021.

Si, Y., and J. P. Reiter (2013), "Nonparametric Bayesian Multiple Imputation for Incomplete Categorical Variables in Large-Scale Assessment Surveys," *Journal of Educational and Behavioral Statistics*, 38, 499–521.

StataCorp. (2021), *Stata Data Analysis Statistical Software: Release 16*, College Station, TX: StataCorp LP, available at http://www.stata.com/. Accessed on September 8, 2021.

Stekhoven, D. J., and P. Bühlmann (2012), "Missforest-Nonparametric Missing Value Imputation for Mixed-Type Data," *Bioinformatics*, 28, 112–118.

Stuart, E. A., M. Azur, C. Frangakis, and P. Leaf (2009), "Multiple Imputation with Large Data Sets: A Case Study of the Children's Mental Health Initiative," *American Journal of Epidemiology*, 169, 1133–1139.

Van Buuren, S. (2012), *Flexible Imputation of Missing Data*, Boca Raton, FL: CRC Press.

Van Buuren, S., and C. Oudshoorn (1999), *Flexible Multivariate Imputation by MICE, Technical Report*, Leiden: TNO Preventie en Gezondheid, TNO/VGZ/PG 99.054.

Welch, C., I. Petersen, J. W. Bartlett, I. R. White, L. Marston, R. W. Morris, I. Nazareth, K. Walters, and J. R. Carpenter (2014), "Evaluation of Two-Fold Fully Conditional Specification Multiple Imputation for Longitudinal Electronic Health Record Data," *Statistics in Medicine*, 33, 3725–3737.

Xie, X., and X. Meng (2017), "Dissecting Multiple Imputation from a Multi-Phase Inference Perspective: What Happens When God's, Imputer's and Analyst's Models Are Uncongenial?," *Statistica Sinica*, 27, 1485–1545.

Xu, D., M. Daniels, and A. G. Winterstein (2016), "Sequential BART for Imputation of Missing Covariates," *Biostatistics*, 17, 589–602.

Zheng, H., and R. J. A. Little (2003), "Penalized Spline Model-Based Estimation of the Finite Populations Total from Probability-Proportional-to-Size Samples," *Journal of Official Statistics*, 19, 99–107.

Zhou, H., M. R. Elliott, and T. E. Raghunathan (2016), "Multiple Imputation in Two-Stage Cluster Samples Using the Weighted Finite Population Bayesian Bootstrap," *Journal of Survey Statistics and Methodology*, 4, 139–170.

Zhu, J., and T. Raghunathan (2015), "Convergence Properties of a Sequential Regression Multiple Imputation Algorithm," *Journal of the American Statistical Association*, 110, 1112–1124.