Imputation in the American Housing Survey: Comparing Multiple Imputation with Current Hot Deck Methods

Sean Dalby, Stephen Ash, Kathy Zha, Gregory Mulley
U.S. Census Bureau, 4600 Silver Hill Road, Suitland, MD 20746

Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Abstract
Multiple imputation is an active field of statistical research, encompassing a wide variety of modeling methods with different strengths and weaknesses. This paper provides a theoretical overview and empirical comparisons between multiple imputation – specifically, Fully Conditional Specification – and traditional hot deck imputation in the context of the American Housing Survey. The hot deck method stratifies across demographic and housing-level characteristics to form donor cells of similar housing units. Unlike the hot deck, Fully Conditional Specification methods allow for a wider array of models, and provide well-researched methods for estimating the amount of variance introduced by imputation itself. Both are compared against hurdles present within the American Housing Survey, including numerous structural zeros, thereby highlighting the benefits and trade-offs of each imputation approach.

Key Words: Bayesian Methods, Multiple Imputation, Hot Deck, Structural Zeros, Fully Conditional Specification, Complex Survey Design

1. Introduction

This paper discusses the development and initial results of multiple imputation (MI) for the American Housing Survey (AHS). The AHS has a number of features that make it an ideal testing ground for MI, such as intricate structural zero dependencies, a wide array of types of variables requiring imputation (continuous, binomial, categorical, etc.), and a complex survey design.

The central goal of this research is to improve upon the current Hot Deck method of imputation within the AHS. The two primary benefits MI brings to bear on data quality are: (1) the ability to measure the variance introduced into estimation by imputation itself and (2) the preservation of complex correlation structures in the dataset at large. These were the two main research goals of this project, though additional practical goals motivated us as well. Specifically, the choice of using Fully Conditional Specification (FCS) to multiply impute missing values allows us to model each variable individually according to their own structural zeros and particular distributions, and more refined modeling with MI ideally produces better point estimates and variances than Hot Deck.

The key results of our research are that MI exceeds Hot Deck in that it improves point estimates, better preserves distributions and correlations, and provides a hitherto unavailable measure of the variance of imputation itself within the AHS. The organization of this paper is as follows: (a) brief overview of the AHS and its current imputation methods, (b) a concise description of MI theory as it pertains to our goals of improving the

current Hot Deck method, (c) a summary of MI diagnostics and comparisons between MI and Hot Deck that support our key results, and finally (d) a conclusion and indications of future potential research.

2. Overview of the American Housing Survey and Its Hot Deck Imputation Method

Before proceeding to the results of our research, this section will introduce the AHS, its sample design, and data processing steps germane to the current imputation setup. It also summarizes the Hot Deck method as it is implemented now, highlighting the aspects needed to compare with the proposed MI route.

2.1 The American Housing Survey

The Department of Housing and Urban Development (HUD) sponsors the most detailed and expansive public survey on housing stock within the United States – the American Housing Survey (AHS) – and works in conjunction with the U.S. Census Bureau to administer this survey. HUD and the Census Bureau collaborated on the AHS sample design and selection, while the Census Bureau is primarily responsible for data collection, editing, and processing, as well as the publication of important sample estimates. As with any survey, AHS has missing data, and it is the Census Bureau's responsibility to impute reasonable values in a well-documented, statistically sound way.

The AHS is a longitudinal survey that selected a new set of housing units (HUs) in 2015. The sample design consists of two main samples: National and Metro. The National AHS employs a complex, two-stage sample design for HUs outside the top 15 largest Core-Based Statistical Areas (CBSAs). The first stage is a stratified sample selection of primary sampling units (PSUs) proportional to size and the second is a systematic sample of housing units within PSUs themselves. The top 15 CBSAs each have 3,000 HUs selected from them systematically, which are combined with the two-stage sample for a total National AHS sample of 84,880 HUs in 2017. This paper deals solely with the 2017 National AHS. For more information, see the Source and Accuracy statement for the 2017 National AHS.

During each two-year survey cycle of the AHS, data processing efforts have to handle two essential types of missing data: unit nonresponse and item nonresponse. Unit nonresponse refers to a noninterview of a HU, so no information is obtained for the HU. The Census Bureau implements a weighting adjustment to account for unit nonresponse. Item nonresponse occurs when an interview is completed, but a few questions remain unanswered. This paper concerns how AHS processing imputes missing values within these types of item nonresponses; unit nonresponse is out of scope.

The "Edits" phase of AHS data processing consists of imputation, blanking, and consistency edits, and occurs after interviewing and before weighting adjustments and disclosure avoidance measures. Edits proceed by module – delimited sets of variables related to each other by theme, such as "Equipment" or "Utilities" modules – and each set of edits for a particular module assume all prior variables in previous modules have been fully edited and imputed. Our research focuses on the first three modules of the AHS: Housing Unit, Out-of-Sequence Household, and Inventory. "Blanking" refers to when certain observations are out of scope for a given variable – such as owners being out of scope for the variable measuring rent amount – and "consistency" refers to particular observations that must be forced to have certain values in order to be consistent with their responses to prior survey questions. For example, a housing unit indicating that it is a

mobile home in a given variable must then be given values consistent with that fact for future related variables. Both blanking and consistency edits exemplify what the MI literature refers to as "structural zeros," where a given observation's responses to prior questions dictate the value it must have for subsequent variables.

2.2 Hot Deck within the AHS

Currently, the Census Bureau employs a form of Hot Deck for the majority of variables requiring imputation in AHS. In the Hot Deck method, groups of variables requiring imputation are assigned specified cells and sort orders based on selected variables, and donors give their observed values to missing observations within the sort order's cells. Proceeding sequentially from the first observation to the last for an imputed variable, the most recent observed value in the sort order serves as the donor for the next missing value(s) until another observed value occurs. The algorithm continues until the next missing value, fills that in with the new, most recently observed value, making sure that donors only give to missing observations within their own sort-order cell. Hot Deck is a relatively simple imputation algorithm, with the added benefit of ensuring viable values for all observations. For our purposes though, it is important to note two things: (1) the sort order provides an implicit model for each variable and (2) frequently, whole groups of variables requiring imputation receive the same implicit model under the existing scheme.

3. Multiple Imputation within the American Housing Survey

MI, properly done, can achieve several statistical enhancements to survey data quality, chief among which include the ability to measure the variance of imputation itself and the preservation of correlations between variables. The choice of using Fully Conditional Specification (FCS) provides further practical benefits in a complex survey with various distributions of variables requiring imputation. This section will provide a short summary of MI theory, highlighting aspects of the theory relevant to these benefits, as well as our specific choices and assumptions we rely on with our research.

3.1 Multiple Imputation: A Non-Technical Overview

Rubin provided the first thorough statistical exposition of MI in 1987 (Rubin, 1987), and the method has since then garnered a significant amount of research attention. The central idea of MI is to impute multiple, varying values in observations that are missing data – yielding multiple datasets – and provide some simple combination rules for these multiple datasets so that a researcher can produce unbiased point estimates that incorporates random variation of imputation itself. Other benefits of MI include, ideally, increased efficiency (Rubin, 1987, p. 16), and preserving correlations in the data.

Anyone imputing data faces several choices and hurdles. The first concern surrounds the mechanism of missingness: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR occurs when the probability of an observation having a missing value in a given variable is uniform throughout the variable (van Buuren, 2019, p. 8). This implies that the missing values in a variable bear no relationship to any variable in the dataset. MCAR is unlikely to occur in practice. MAR, by contrast, means that the probability of missingness for a given variable is uniform *after* conditioning upon other variables in the dataset. MNAR occurs when there is a pattern to the missingness in a variable that is unexplained by observed variables.

The second decision surrounds the particular choice of imputation method – the actual "imputation" portion of MI theory. Van Buuren (2019; p. 67-91) gives a concise

overview of potential imputation methods, some of which include classic regression prediction (including stochastic regression, which attempts to draw from an estimated distribution), predictive mean matching, classification trees, and so on. All of these methods for actually selecting an imputed value are compatible with MI.

Currently, the two major routes of MI are Joint-Modeling and FCS. Joint-Modeling calls for a precise specification of the joint distribution of all variables requiring imputation (Akande et al., 2018). The joint distribution's parameters are then estimated using observed data, and multiple values are independently drawn from the estimated distribution. This route is elegant in theory, but fairly intractable in practice. For example, it is extremely difficult, if not impossible, to specify a multivariate distribution containing a combination of continuous, binomial, categorical, and Poisson variables.

By contrast, FCS – also called Sequential Regressions (Raghunathan, 2016) – imputes one variable at a time according to the appropriate regression model of that variable and, after imputing every variable necessary in the data, iterates back over the set of variables for a new set of imputations until the distributions stabilize. This iterative process allows imputed variables to use imputed values of other variables within each iteration beyond the first. FCS is flexible enough to handle any common distribution, it can work with complex structural zero dependencies, and it allows for a wide range of modeling choices for each individual variable.

Under FCS, the imputer needs to consider (1) how many burn-in iterations should there be, (2) what order (if any) should the variables be imputed in, (3) how many imputed datasets should be retained, (4) do the conditional models stem from the same joint distribution, and finally (5) how will the particular models compare with analysts' models. Estimates generally stabilize quickly, depending on the amount of missing data in a given variable (Raghunathan, 2016, p. 69); visual diagnostics assist in determining this burn-in point. Raghunathan addresses the order question, and concludes that order is essentially unimportant (Raghunathan, 2016, p. 68). Five is a standard choice of number of imputed datasets to use in the diagnostic phase, though this should ideally be higher for production purposes (van Buuren, 2019).

The fourth and fifth questions above concern "compatibility" and "congeniality," respectively. Technically, FCS is only valid if the conditional distributions of each regression are compatible, meaning they have the same joint distribution. If no such joint distribution exists, the underlying Gibbs Sampler mechanism is not guaranteed to converge (Li et al., 2012). Some contemporary research indicates that FCS is reasonably robust against deviations from this assumption (van Buuren et al., 2006; Zhu and Raghunathan, 2015). The seminal work of Meng (1994) demonstrated the importance of congeniality as well; the regression model used to impute data must be equivalent – or broader in scope – than an analyst's model. Otherwise, point estimates of analysts will be biased (van Buuren, 2019).

Finally, we should give some thoughts to incorporating the National AHS' complex survey design into our imputation methodology. Reiter et al. (2006) suggests including strata or cluster dummy variables within models to account for design features, but acknowledge there appear to be cases where design variables have no meaningful effect on imputation.

These are the main aspects of MI theory relevant to our work. We need to address the missingness mechanism, choose an imputation procedure, pick between Joint-Modeling and FCS, make familiar modeling choices, pick a number of iterations and datasets, consider the theoretical validity of our methods with respect to compatibility and congeniality, as well as the impact of a complex survey design.

3.2 Assumptions and Decisions for MI within the AHS

We have considered each of the above problems in our MI research. However, due to the ongoing nature of our research into MI within the AHS, some of our decisions are temporary stopgap measures based on best practices or widely accepted assumptions. We will elaborate on our decision process for each question below, highlighting when we assumed something and when we chose something based on empirical results.

With regard to the missingness mechanism, this study assumes all variables are MAR. There are no commonly accepted tests for identifying a missing data mechanism, though some current research is being done in this area (Tshering et al., 2013). MNAR is particularly difficult to test, since, by definition, MNAR occurs when there are unobserved influences on missingness. Still, some of our results discussed in Section 4 suggest either that (1) some variables need further refined models to ensure MAR or (2) there is an unknown factor worth pursuing with subject-matter experts.

We chose to use stochastic regression modeling with randomness inserted into both the predictions and the estimated regression coefficient parameters as our imputation method. This is what van Buuren refers to as "Bayesian MI" (van Buuren, 2019, p. 67), since it considers the regression coefficient parameters to have distributions themselves. We also chose to use FCS, rather than Joint-Modeling, for the "multiple" part of MI theory. There is evidence suggesting FCS is better at imputing categorical variables (Kropko et al., 2013), of which there are many in the AHS. Given this, the prevalence of complicated structural zero dependencies within the AHS, and the wide range of distributions present across the set of variables requiring imputation, FCS seemed by far the logical choice. Our technical method for drawing from the posterior distribution of coefficients and predicted values mimics the processes described by Raghunathan (2016, p. 69-73) and van Buuren (2019, p. 68, 88).

We kept our first fully imputed dataset at iteration ten and took every tenth dataset after that until iteration 50, for a total of five datasets used to calculate point estimates and variances. We chose five datasets since this is standard for diagnostic purposes (van Buuren, 2019, p. 340), and we chose the tenth iteration in part on a general rule of thumb (van Buuren, 2019, p. 120) and in part on the convergence of means we found in our variables. We chose a gap of ten between selecting datasets to ensure independence.

The issues of compatibility and congeniality are considerably more difficult to assess. We assume our conditional distributions have the same joint-distribution, or at least that any deviations from this have minimal impact on our results. The literature cited above bolsters this assumption by suggesting FCS is robust to violations of compatibility. Since the AHS is a general-purpose survey, it is impossible to build an imputation model encompassing all conceivable models any data analyst will use with our data. Therefore, we follow the advice in Murray (2018) and try to mitigate congeniality issues by incorporating many variables into our regression models within modest computational constraints (Murray, 2019, p.9). Our general modeling method used a standard forward-selection algorithm using the Akaike Information Criterion (AIC) as the selection

criterion, which considered every possible variable within the given structural zero domain as potential independent variables. The set of variables considered for modeling purposes included every variable within the three modules under consideration, some demographic AHS variables, and a small set of variables taken from the American Community Survey (ACS) on the Census Bureau's Planning Database.

Our models thus far only include one design variable: a binary indicator of being in the top 15 CBSAs. Future model refinements will likely include additional design variables, though our understanding of current research suggests that design variables should be included in an abundance of caution, not because they are vital for every variable's model in every case. The conclusion of Reiter et al. (2006), for example, is essentially: design variables should be included in a model when they are correlated with the outcome of interest. The same could be said of any non-design variable. Granted, design variables are frequently correlated with outcomes – which is why we agree with the abundance of caution approach – but, again, so are many demographic and other variables deemed generally important by subject-area experts. Overall then, we encourage the use of design variables in modeling, even if their unique role in MI remains somewhat opaque to us.

4. MI Diagnostics and Hot Deck Comparisons

This section first describes the rate of missingness for the imputed variables under consideration here, and then turns to diagnostic output. The empirical results of this research take two main forms: (1) assessing the quality of the models and MI methods themselves and (2) comparing MI with the current Hot Deck method. While there has been some recent research on more sophisticated diagnostic tools (Nguyen et al., 2017), standard MI diagnostics evaluate the convergence of means and standard deviations across iterations, compare observed and imputed distributions, typically via a validation study, and use traditional model fitting and distributional discrepancy tools (van Buuren, 2019, p. 51, 187, 190). We examined the convergence of means and standard deviations and compared observed verses imputed distributions in a validation study. Our Hot Deck comparisons include distribution and point estimate comparisons of several different types of variables, as well as some tables contrasting correlations preserved between the methods.

4.1 Rates of Missingness

Unsurprisingly, variables requiring imputation within the first three modules of the AHS have varying amounts of missingness within them. These range from near zero percent to 42 percent missing, with multinomial and binomial variables generally having lower rates of missingness than continuous and count (modeled as Poisson distributions) variables. Table 1 below gives a sense of the amount of missing data throughout our imputed variables. The first column describes the given variable, and the second states its mathematical distribution. The third gives the number of valid observations within the variable's domain, and the fourth lists the percentage of missing data for each given variable.

Table 1: Rates of Missingness among Imputed Variables

Variable	*Distribution*	*Rate of Missingness*
Type of Building	Categorical	0.09%
No. of Units	Continuous	3.05%
Tenure	Categorical	0.02%
Entry System for HU	Binomial	10.17%
HU in Apartment Complex	Binomial	0.80%
Type of HU	Categorical	0.10%
Type of Mobile Home	Categorical	0.33%
Type of Vacancy	Categorical	9.96%
Monthly Rent Amount	Continuous	10.85%
No. of Full Bathrooms	Count	1.10%
No. of Bedrooms	Count	1.25%
Year Built	Continuous	1.56%
No. of Dens	Count	34.59%
No. of Dining Rooms	Count	1.72%
No. of Family Rooms	Count	30.41%
No. of Half Bathrooms	Count	1.47%
No. of Kitchens	Count	1.05%
No. of Laundry Rooms	Count	17.00%
No. of Living Rooms	Count	1.24%
No. of Other Finished Rooms	Count	37.68%
No. of Unfinished Rooms	Count	42.53%
No. of Recreation Rooms	Count	40.26%
Market Value of HU	Continuous	15.00%
Routine Maintenance Cost	Continuous	11.27%
HU Made Accessible	Binomial	41.93%
HU Made Energy Efficient	Binomial	41.99%
HU Made Ready for Sale	Binomial	41.94%
Is Anchored	Binomial	5.76%
Has a Basement	Categorical	0.68%
Is a Condo / Cooperative	Categorical	0.52%
No. of Floors	Count	0.56%
Has a Garage	Binomial	0.51%
Gut Rehabilitation	Binomial	1.39%
Square Footage of Lot	Continuous	2.34%
Mobile Home Foundation	Categorical	1.58%
No Entrance Steps	Binomial	0.40%
Has a Porch	Binomial	0.58%
No. of Stories	Count	0.56%
No. of Mobile Homes	Continuous	8.48%
Unit Square Footage	Continuous	11.84%

Source: U.S. Census Bureau, 2017 American Housing Survey

4.2 MI Diagnostics

4.2.1 Mean and Standard Deviation Convergence

Overall, means and standard deviations for our FCS method converged more quickly for categorical and binomial variables than for continuous and count variables. This appears to be independent of amount of missingness in each type of variable. Multinomial and binomial variables converge almost immediately, while some continuous and count variables take as long as iteration 30 to converge. Figure 1a below show the mean and standard deviation convergence for a categorical variable describing the particular vacancy status of a vacant housing unit, and Figure 1b display analogous results for a binomial variable indicating whether a certain home renovation was done to increase energy efficiency within the housing unit.

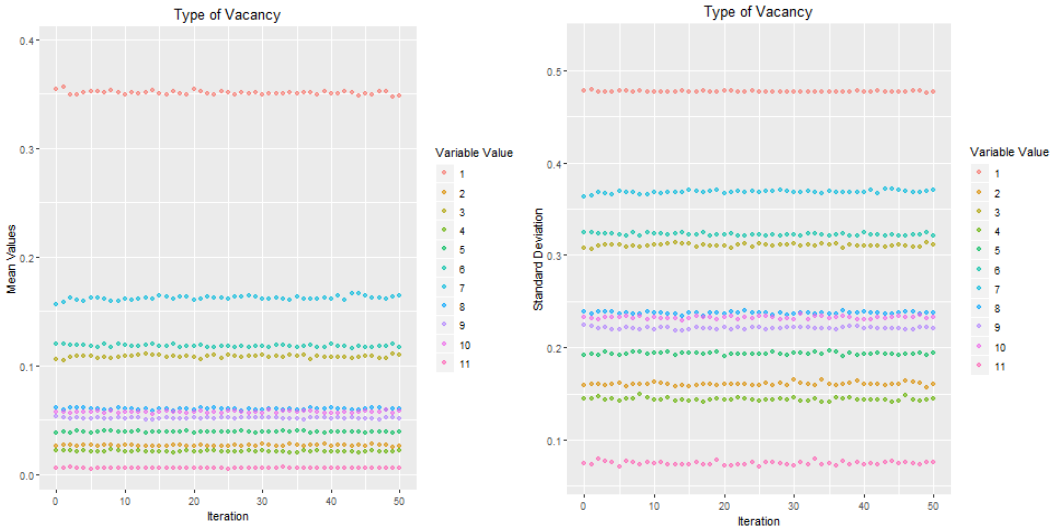

Figure 1a: Type of vacancy – mean / standard deviation convergence. Missingness rate: 1.3%. Source: U.S. Census Bureau, 2017 American Housing Survey

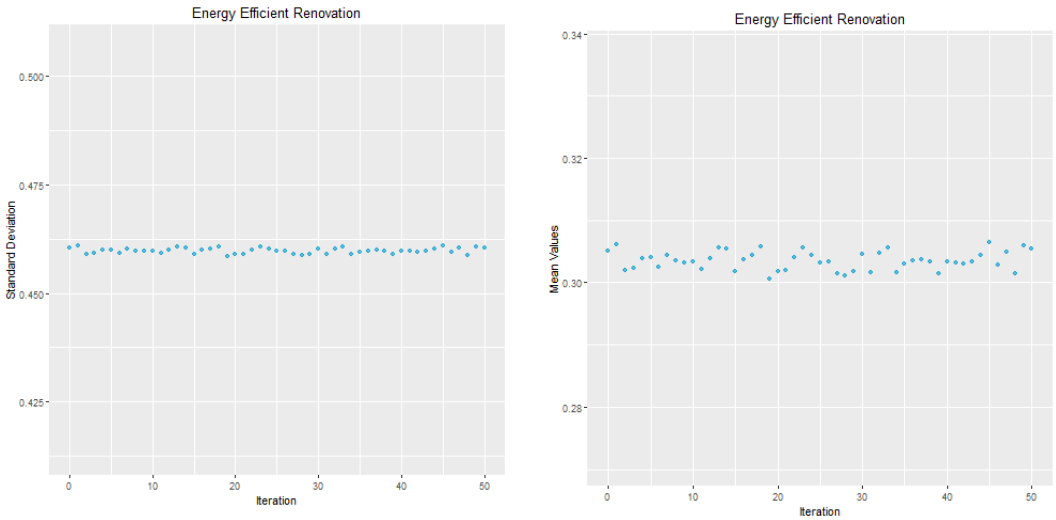

Figure 1b: Efficient renovation – mean / standard deviation convergence. Missingness rate: 21.1%. Source: U.S. Census Bureau, 2017 American Housing Survey

The figures above stand in stark contrast to the continuous and count variables below. Figures 1c/d depict a continuous variable asking about the value of the housing unit and a count variable asking how many utility or laundry rooms are in the housing unit, respectively.

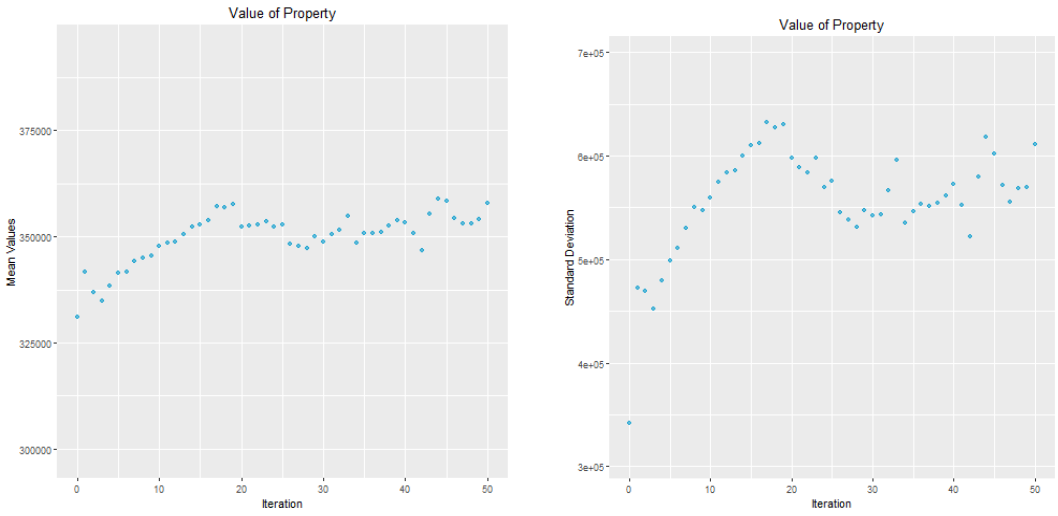

Figure 1c: Value – mean / standard deviation convergence. Missingness rate: 8.9% Source: U.S. Census Bureau, 2017 American Housing Survey

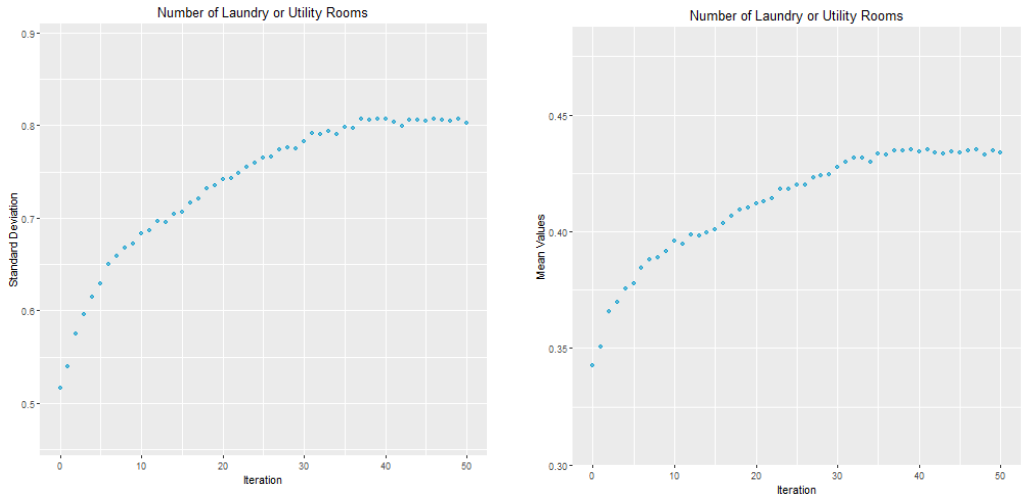

Figure 1d: Laundry rooms – mean / standard deviation convergence. Missingness rate: 17% Source: U.S. Census Bureau, 2017 American Housing Survey

Not all continuous or count variables take this long to converge – the variable for rent only takes about 4 iterations to converge – but they indicate that we may want to consider longer burn-in iterations before selecting imputed datasets for the AHS. Generally though, the fact that all of these variables converge indicates a well-functioning FCS MI implementation or, more specifically, that there exists stable conditional distributions to which this algorithm converges.

4.2.2 Validation Study

The value of a validation study centers around our ability to control the mechanism of missingness. Our validation study inserts missing values randomly into ten percent of observed data (thereby ensuring the missing data are MCAR), applies our FCS MI implementation on the simulated missing data, and compares the distribution of actual verses imputed data. In this way, no real nonresponse observations affected this analysis, and we are able to see how close our imputation methods approximate observed values.

The comparison for categorical, binomial, and count variables takes the form of two kinds of histograms: one depicting the frequency of each discrete value in the real and imputed artificially missing data, and another showing the distribution of differences between real and imputed values for each observation in the artificially missing data. Continuous variables use a scatter plot and a histogram, with real values along the x-axis and imputed values along the y-axis for the scatter plot, and the differences shown again in the histogram. We need both plots to confirm that distributions of imputed data reasonably mirror real data. The frequency count and scatter plot confirm the distributions are similar, and high counts at (or near) zero in the differences histogram indicate that accurate values are being imputed in individual observations.

Figures 2a/b/c/d/e below indicate the well-functioning imputation models used in the four variables used above (a categorical, binomial, count, and continuous variable, respectively), as well as another count variable – number of bedrooms – to illustrate the varying degrees of success with the count variables.

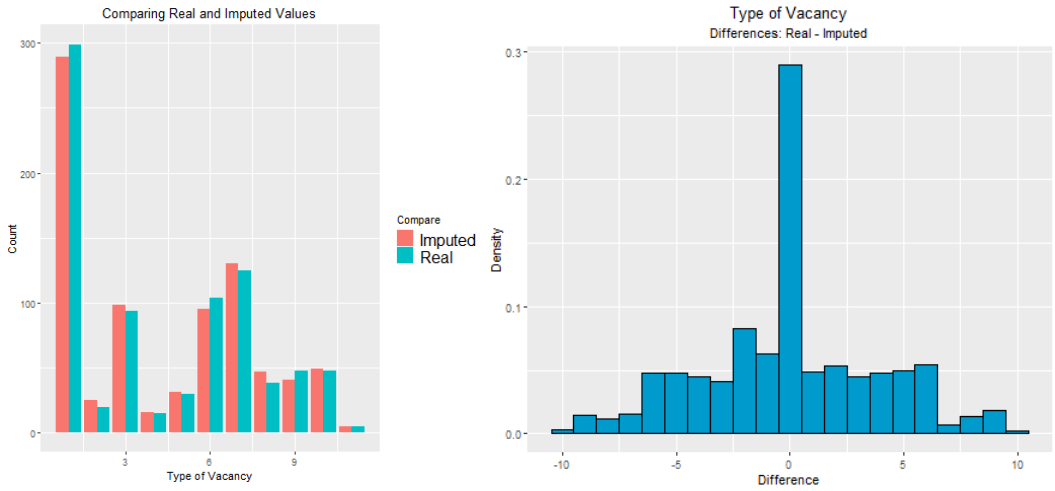

Figure 2a: Type of vacancy – validation comparison. Distribution: Categorical
Source: U.S. Census Bureau, 2017 American Housing Survey

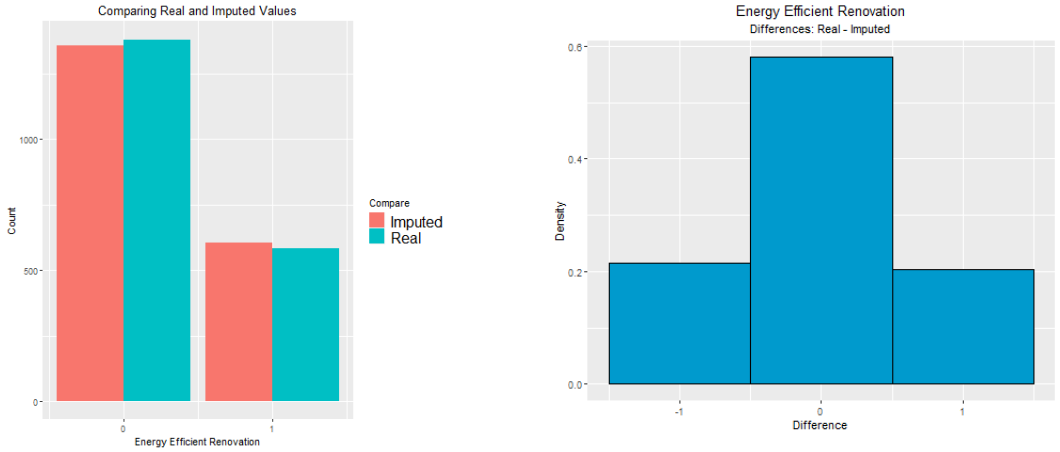

Figure 2b: Efficient renovation – validation comparison. Distribution: Binomial
Source: U.S. Census Bureau, 2017 American Housing Survey

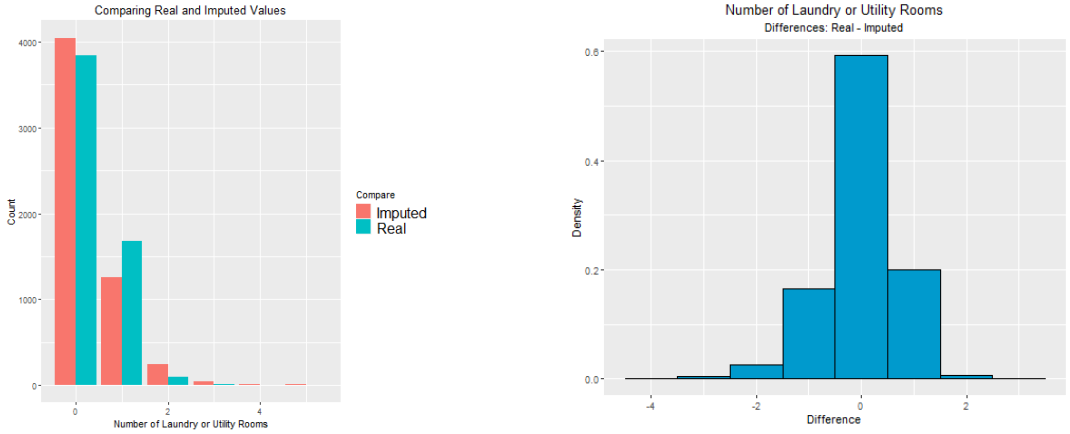

Figure 2c: Number of laundry rooms – validation comparison. Distribution: Poisson
Source: U.S. Census Bureau, 2017 American Housing Survey

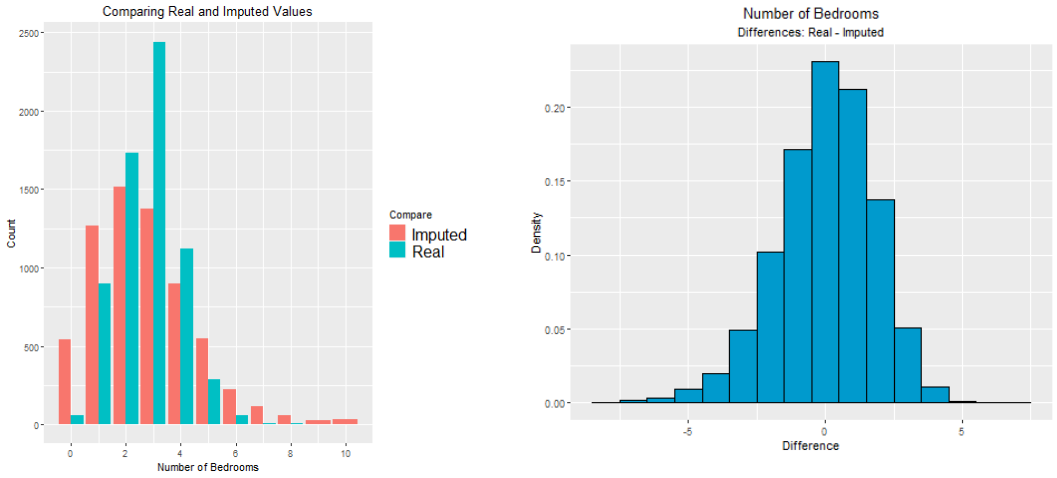

Figure 2d: Number of bedrooms – validation comparison. Distribution: Poisson
Source: U.S. Census Bureau, 2017 American Housing Survey

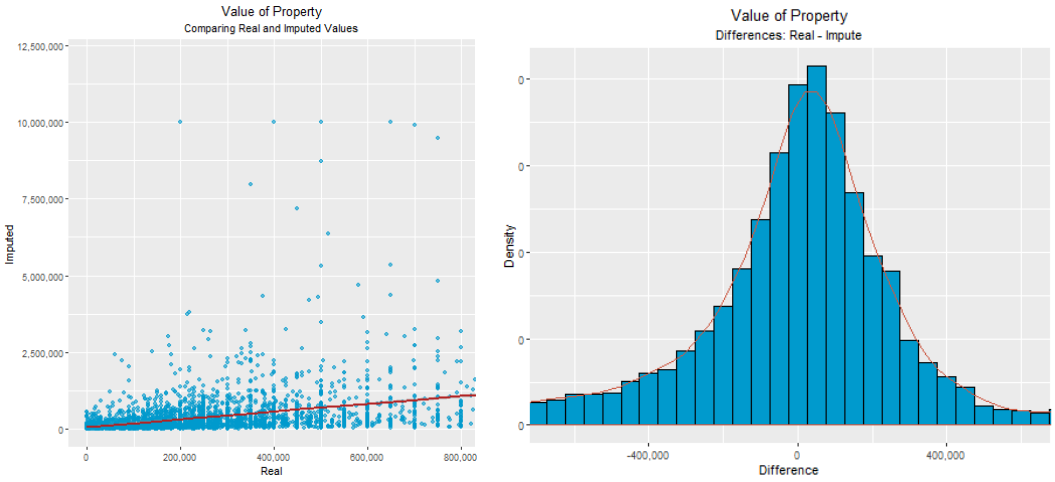

Figure 2e: Value – validation comparison. Distribution: Normal / Continuous
Source: U.S. Census Bureau, 2017 American Housing Survey

Again, categorical and binomial variables do extremely well with these diagnostic measures. Count variables fair somewhat well, and continuous variables suffer from some over-estimation in certain cases. As Figure 2e suggests, the current MI algorithm imputes some significant outliers in some continuous variables and Figure 2d indicates that the variable counting bedrooms in a HU also diverges somewhat from the real distribution. These results suggest that we should focus future research on model refinements. Still, the validation study strongly indicates the viability of MI within the AHS, despite the need for ongoing model tweaking.

Overall, we are reasonably confident in the performance of our MI algorithm. It converges well and produces reasonable imputations when compared against known values under MCAR, while acknowledging further model refinements and diagnostics need to be pursued for certain variables. In particular, this validation simulation should be expanded to include examining correlations, varying the types of missingness mechanisms, looking at averages over simulations, applying and comparing results to the Hot Deck method, and varying the amount of randomness inserted into the variables.

4.3 MI vs. Hot Deck in the AHS

Now, we turn to comparing MI to the Hot Deck. This section focuses on three central points supporting this conclusion: MI improves point estimates, and it better preserves distributions and correlations. Furthermore, unlike Hot Deck, MI provides an estimation of the variance of imputation itself, which is an automatic improvement over Hot Deck.

4.3.1 Point Estimates and Imputed Variance

MI should provide more accurate point estimates than the current Hot Deck imputation method for both practical and theoretical reasons. Practically speaking, the current Hot Deck cells and sort orders in the AHS are outdated, and frequently use one implicit model for multiple variables at a time. This means that any attempts to improve on these will likely succeed. More theoretically though, Hot Deck methods do not incorporate the full range of information provided by continuous variables, since these variables have to be binned in order to use them in a sort order. Our MI implementation of FCS uses common

regression modeling as its foundation, and as such can use the full amount of information of any variable within its models.

The Fraction of Missing Information (FMI) is the percent of total variance due to the imputation process. This statistic the ratio of the imputation variance of the variable and the total variance with some additional correcting terms depending on the number of imputed datasets (Enders, 2010, p. 222-226). If the number of imputed datasets is high, then FMI is approximated by equation (1) below, where m is the number of imputed datasets, V_B is the variance between imputed datasets and V_T is the total variance:

$$FMI = \frac{V_B + {}^{V_B}/_{m}}{V_T} \qquad (1)$$

Since we used only five datasets, formula (2) below is more appropriate for our analysis, where v is a degrees of freedom term and V_W is the average variance within each imputed dataset:

$$FMI = \frac{V_B + {}^{V_B}/_{m} + {}^{2V_W}/_{(v+3)}}{V_T} \qquad (2)$$

Higher percentages of FMI suggest a greater need for MI methods for a given variable because, absent a measure of imputation variance, confidence intervals for parameter estimates of a variable with high FMI will be egregiously narrow. A higher FMI also suggests that a given variable will take longer to converge (Enders, 2010, p. 226), so it serves as a good diagnostic tool as well.

Our models employed in our MI algorithm improve upon the existing Hot Deck results. To illustrate these improvements, Tables 2a/b/c/d below provide the weighted means of many of the variables in the three AHS modules we have imputed with MI. Table 2a displays the proportions of binomial variables, 2b has proportions for categorical variables, 2c has means for count variables, and 2d has means for continuous variables. The second column – "Observed" – is the mean of the given variable based on the observed data, the third column – "Hot Deck" – is the mean based on the Hot Deck, and the fourth column – "MI" – is the mean based on MI. The fifth column is FMI. Green rows occur when the MI estimate is closer to the observed estimate than Hot Deck, yellow implies that the MI and Hot Deck means are extremely close[1] to the observed value, and red means Hot Deck performs better than MI.

[1] "Extremely close" here just means within one order of magnitude for a particular distribution. The categorical and binomial variables need to be within 0.01 (or one percent) of the observed proportion, the mean of imputed count variables within 0.1 of the observed count mean, and imputed means for continuous variables within one unit of the observed mean.

Table 2a: FMI / Proportions – Binomial

Variable	Observed	Hot Deck	MI	FMI
Entry System for HU	0.396	0.407	0.398	3.68%
Gut Rehabilitation	0.201	0.201	0.201	2.30%
Has a Garage	0.635	0.635	0.635	0.61%
Has a Porch	0.843	0.843	0.843	0.69%
HU in Apartment Complex	0.610	0.609	0.609	0.62%
HU Made Accessible	0.066	0.067	0.066	45.99%
HU Made Energy Efficient	0.302	0.308	0.303	26.25%
HU Made Ready for Sale	0.036	0.038	0.035	56.86%
Is Anchored	0.859	0.865	0.865	5.47%
No Entrance Steps	0.504	0.504	0.503	0.11%

Table 2b: FMI / proportions – Categorical

Variable	Observed	Hot Deck	MI	FMI
Basement Under Part of HU	0.100	0.104	0.105	0.30%
Basement Under Whole HU	0.298	0.307	0.308	0.41%
Concrete Slab Foundation	0.370	0.352	0.350	0.21%
Crawl Space	0.209	0.215	0.215	0.28%
For Rent Only	0.301	0.288	0.299	9.26%
For Rent or for Sale	0.025	0.023	0.026	14.19%
For Sale Only	0.109	0.108	0.111	16.11%
Held for Occasional Use	0.136	0.153	0.134	8.35%
Is a Co-Op	0.008	0.008	0.008	0.29%
Is a Condo	0.065	0.064	0.065	0.72%
Migratory HU	0.006	0.005	0.006	14.50%
Neither a Condo nor a Co-Op	0.923	0.928	0.923	0.68%
Other Cellar	0.023	0.022	0.022	0.08%
Other Vacancy	0.175	0.199	0.178	17.46%
Owner Occupied	0.640	0.638	0.638	0.01%
Rented, but not Occupied	0.016	0.016	0.016	0.00%
Rented, but not Occupied	0.021	0.019	0.021	5.32%
Renter Occupied	0.343	0.346	0.346	0.01%
Seasonal-Other	0.066	0.060	0.066	6.36%
Seasonal-Summer	0.082	0.074	0.081	2.87%
Seasonal-Winter	0.041	0.037	0.040	1.09%
Sold, but not Occupied	0.037	0.033	0.037	7.11%

Table 2c: FMI / means – Count

Variable	Observed	Hot Deck	MI	FMI
No. of Bedrooms	2.760	2.758	2.775	66.46%
No. of Dens	0.159	0.100	0.304	98.77%
No. of Dining Rooms	0.494	0.492	0.503	63.69%
No. of Family Rooms	0.215	0.143	0.241	76.22%
No. of Floors	1.813	1.790	1.770	0.06%
No. of Full Bathrooms	1.681	1.679	1.698	86.45%
No. of Half Bathrooms	0.317	0.317	0.337	91.15%
No. of Kitchens	1.009	1.009	1.009	27.22%
No. of Laundry Rooms	0.377	0.311	0.459	97.82%
No. of Living Rooms	1.044	1.044	1.049	47.78%
No. of Other Finished Rooms	0.139	0.084	0.205	94.54%
No. of Recreation Rooms	0.069	0.039	0.110	93.27%
No. of Stories	1.813	2.225	1.771	0.04%
No. of Unfinished Rooms	0.035	0.019	0.100	97.91%

Table 2d: FMI / means – Continuous

Variable	Observed	Hot Deck	MI	FMI
Market Value of HU	292,400	299,100	311,600	67.58%
Monthly Rent	1,010	1,072	1,048	59.01%
No. of Mobile Homes	41	45	41	6.48%
No. of Units	10.38	11.35	11.21	5.67%
Routine Maintenance Costs	801	865	871	23.23%
Unit Square Footage	1,654	1,703	1,796	96.97%
Year Built	1970	1970	1970	0.60%

Tables 2a/b indicate that MI rivals or exceeds Hot Deck on most measures, even where there is a substantial FMI. We can see too that the variables taking somewhat longer to converge – the rent and number of bedrooms variables – indeed have a higher FMI.

4.3.2 Preserving Distributions

Under a MAR assumption, any imputation method should replicate the distribution of known values successfully (Raghunathan, 2016, p. 87). To test this, we can compare what the distributions of imputed values under Hot Deck and MI look like against the observed distribution of values. Figures 3a/b/c below all show how MI performs better than Hot Deck method in replicating the distribution of observed values.

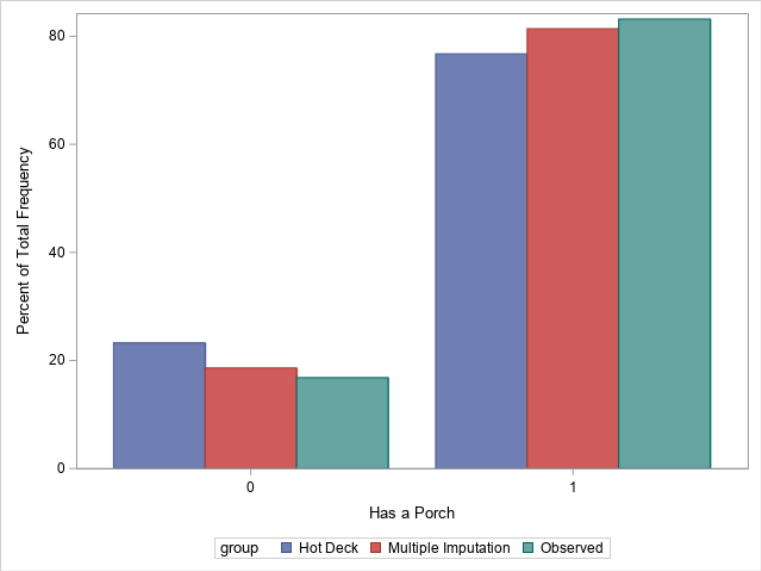

Figure 3a: Comparing Distributions – Has a Porch
Source: U.S. Census Bureau, 2017 American Housing Survey

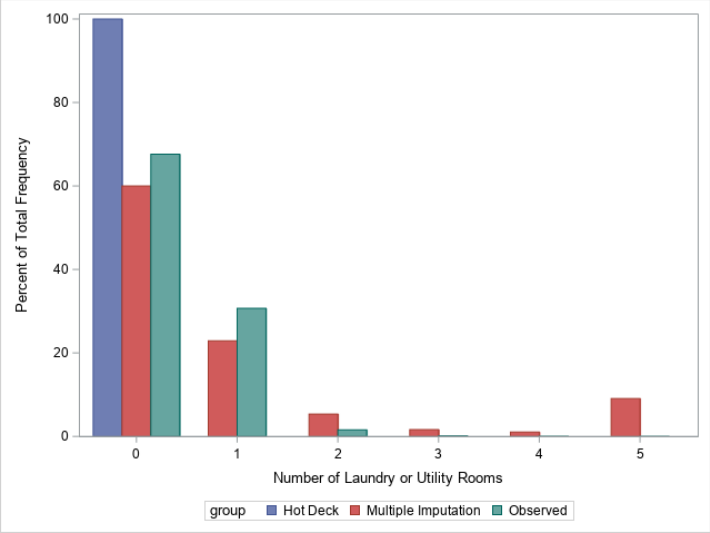

Figure 3b: Comparing Distributions – Number of Laundry Rooms
Source: U.S. Census Bureau, 2017 American Housing Survey

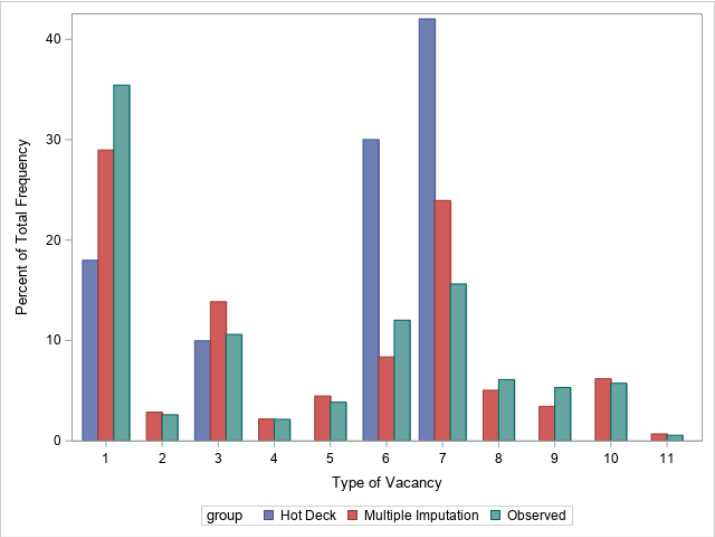

Figure 3c: Comparing Distributions – Type of Vacancy
Source: U.S. Census Bureau, 2017 American Housing Survey

With a few exceptions, MI is closer to the actual frequency of observed values in these variables than Hot Deck. The laundry variable in Figure 3b has a curious spike at 5 under MI, which it did not have during the validation study. This again suggests more model refinement for this variable, but it highlights the need for multiple diagnostic measures when assessing MI. We would not have caught this merely by looking at convergence rates and the validation results. Still, these comparisons highlight another potential benefit of MI within the AHS.

4.3.3 Preserving Correlations

Successful implementations of MI can preserve correlations between variables when they are incorporated into each others' models. Since MI allows – even encourages – extremely broad imputation models, this theoretical result implies that most variables' correlations can be maintained easily with a wide enough scope in modeling.

To measure this within this study, we constructed a relatively intuitive correlation-adjacent statistic. We calculated the correlation of binomial and continuous variables under consideration in our study with just the observed data, then with the observed plus Hot Deck imputations, and then with the observed and MI imputed values. We then made two tables: one of the absolute value of the difference between the observed and Hot Deck, and another of the absolute value of the difference between the observed and MI. Two continuous or two binomial variables' correlation was measured using standard Pearson correlations, and the correlation between a continuous and binomial variable was measured as Point-Serial correlations. The statistic takes the form of formulas (3) and (4) below, where "HD" means "Hot Deck," "MI" means "Multiple Imputation," "Obs" means "Observed" and *r* indicates the relevant correlation:

$$\left| r_{HD,\, var1,\, var2} - r_{Obs,\, var1,\, var2} \right| \tag{3}$$

$$\left| r_{MI,\, var1,\, var2} - r_{Obs,\, var1,\, var2} \right| \tag{4}$$

Tables 3a/b below show the results from these correlation comparisons. Green indicates an absolute difference close to zero, meaning that the correlations before and after the given imputation method preserved the correlations among the observed values. Shades towards red indicate more divergence from the initial correlation.

Table 3a: Hot Deck and Observed Difference of Correlations

Variable	RENT	BUILT	VALUE	CSTMNT	UNITSF	LOT	GARAGE	ANCHOR	NOSTEP	GUTREHB	HMRACCE	HMRENEF	HMRSALE	ACCESSB	COMPLEX	PORCH
RENT	0.000	0.048			0.182	0.061	0.119	0.019	0.011					0.066	0.018	0.030
BUILT	0.048	0.000	0.010	0.000	0.089	0.042	0.001	0.019	0.002	0.001	0.001	0.000	0.001	0.000	0.004	0.001
VALUE		0.010	0.000	0.027	0.202	0.021	0.064	0.050	0.003	0.002	0.009	0.001	0.002	0.031	0.079	0.026
CSTMNT		0.000	0.027	0.000	0.034	0.005	0.006	0.016	0.003	0.002	0.000	0.001	0.007	0.008	0.021	0.007
UNITSF	0.182	0.089	0.202	0.034	0.000	0.017	0.196	0.032	0.043	0.015	0.012	0.000	0.009	0.014	0.021	0.126
LOT	0.061	0.042	0.021	0.005	0.017	0.000	0.024	0.032	0.037	0.016	0.014	0.004	0.009			0.033
GARAGE	0.119	0.001	0.064	0.006	0.196	0.024	0.000	0.009	0.000	0.000	0.001	0.001	0.000	0.010	0.003	0.001
ANCHOR	0.019	0.019	0.050	0.016	0.032	0.032	0.009	0.000	0.002	0.009	0.003	0.010	0.004			0.016
NOSTEP	0.011	0.002	0.003	0.003	0.043	0.037	0.000	0.002	0.000	0.000	0.001	0.001	0.000	0.007	0.001	0.000
GUTREHB		0.001	0.002	0.002	0.015	0.016	0.000	0.009	0.000	0.000	0.002	0.002	0.001	0.007	0.001	0.000
HMRACC		0.001	0.009	0.000	0.012	0.014	0.001	0.003	0.001	0.002	0.000	0.003	0.001	0.005	0.009	0.002
HMRENE		0.000	0.001	0.001	0.000	0.004	0.001	0.010	0.001	0.002	0.003	0.000	0.000	0.001	0.009	0.001
HMRSALE		0.001	0.002	0.007	0.009	0.009	0.000	0.004	0.000	0.001	0.001	0.000	0.000	0.004	0.002	0.001
ACCESSB	0.066	0.000	0.031	0.008	0.014		0.010		0.007	0.007	0.005	0.001	0.004	0.000	0.001	0.007
COMPLEX	0.018	0.004	0.079	0.021	0.021		0.003		0.001	0.001	0.009	0.009	0.002	0.001	0.000	0.004
PORCH	0.030	0.001	0.026	0.007	0.126	0.033	0.001	0.016	0.000	0.000	0.002	0.001	0.001	0.007	0.004	0.000

Source: U.S. Census Bureau, 2017 American Housing Survey

Table 3b: MI and Observed Difference of Correlations

Variable	RENT	BUILT	VALUE	CSTMNT	UNITSF	LOT	GARAGE	ANCHOR	NOSTEP	GUTREHB	HMRACCE	HMRENEF	HMRSALE	ACCESSB	COMPLEX	PORCH
RENT	0.000	0.022			0.183	0.028	0.048	0.051	0.007					0.033	0.022	0.004
BUILT	0.022	0.000	0.013	0.001	0.154	0.010	0.001	0.013	0.000	0.002	0.008	0.006	0.010	0.001	0.007	0.000
VALUE		0.013	0.000	0.024	0.208	0.010	0.038	0.021	0.014	0.013	0.019	0.012	0.001	0.055	0.102	0.001
CSTMNT		0.001	0.024	0.000	0.006	0.016	0.017	0.015	0.007	0.015	0.004	0.009	0.009	0.026	0.011	0.009
UNITSF	0.183	0.154	0.208	0.006	0.000	0.003	0.283	0.084	0.070	0.015	0.020	0.016	0.011	0.013	0.019	0.170
LOT	0.028	0.010	0.010	0.016	0.003	0.000	0.005	0.048	0.001	0.011	0.010	0.000	0.001			0.013
GARAGE	0.048	0.001	0.038	0.017	0.283	0.005	0.000	0.014	0.001	0.001	0.007	0.004	0.003	0.002	0.001	0.000
ANCHOR	0.051	0.013	0.021	0.015	0.084	0.048	0.014	0.000	0.006	0.015	0.013	0.001	0.004			0.008
NOSTEP	0.007	0.000	0.014	0.007	0.070	0.001	0.001	0.006	0.000	0.001	0.005	0.002	0.006	0.007	0.001	0.000
GUTREHB		0.002	0.013	0.015	0.015	0.011	0.001	0.015	0.001	0.000	0.023	0.033	0.009	0.018	0.003	0.001
HMRACC		0.008	0.019	0.004	0.020	0.010	0.007	0.013	0.005	0.023	0.000	0.007	0.006	0.033	0.001	0.004
HMRENE		0.006	0.012	0.009	0.016	0.000	0.004	0.001	0.002	0.033	0.007	0.000	0.008	0.003	0.002	0.010
HMRSALE		0.010	0.001	0.009	0.011	0.001	0.003	0.004	0.006	0.009	0.006	0.008	0.000	0.010	0.048	0.002
ACCESSB	0.033	0.001	0.055	0.026	0.013		0.002		0.007	0.018	0.033	0.003	0.010	0.000	0.005	0.006
COMPLEX	0.022	0.007	0.102	0.011	0.019		0.001		0.001	0.003	0.001	0.002	0.048	0.005	0.000	0.001
PORCH	0.004	0.000	0.001	0.009	0.170	0.013	0.000	0.008	0.000	0.001	0.004	0.010	0.002	0.006	0.001	0.000

Source: U.S. Census Bureau, 2017 American Housing Survey

We can see that there are a handful of binomial variables with slightly varying correlational structures after MI, and some variables remain somewhat divergent under both Hot Deck and MI. However, the lot size variable shows marked improvement under MI, and the overall replication of correlations is supportive of MI. The flexibility of future model refinements gives us more confidence in the benefit of MI over Hot Deck on this measure as well. For example, we can easily add more variables to binomial models to improve preserved correlations, whereas it is theoretically unclear that doing so with Hot Deck would improve anything.

5. Conclusion and Future Research

This research strongly suggest the viability of MI for the AHS, though future research should focus on model refinement for key variables. The MI algorithm we developed converges properly and performs reasonably well in our validation study. More importantly, MI improves upon point estimates, better preserves distributions and correlations, and provides a measure of imputation variance itself for the AHS. Future research certainly includes more model development, but should also include revisiting the missing data mechanism assumption, as well as the number of iterations required for convergence and number of complete datasets.

References

Akande, O, Reiter, J. & Barrientos, A. (2018). Multiple Imputation of Missing Values in Household Data with Structural Zeros. *Survey Methodology*.

Enders, C. (2010). *Applied Missing Data Analysis*. New York: The Guilford Press.

Kropko, J., Goodrich, B., Gelman, A., & Hill, J. (n.d.). (2013). Multiple Imputation for Continuous and Categorical Data: Comparing Joint Multivariate Normal and Conditional Approaches. *Political Analysis, 22*(4), 497-519.

Li, F., Yu, Y. & Rubin, D. (2012). Imputing missing data by fully conditional models: Some cautionary examples and guidelines. *Duke University Department of Statistical Science Discussion Paper*, 11-24.Murray, J. (2018). Multiple Imputation: A Review of Practical and Theoretical Findings. *Statistical Science, 33*(2), 142-159.

Murray, J. (2018). Multiple Imputation: A Review of Practical and Theoretical Findings. *Statistical Science, 33*(2), 142-159.

Meng, X. (1994). Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science, 9*(4), 538-573.

Nguyen, C., Carlin, J., Lee, K. (2017). Model checking in Multiple Imputation: an overview and case study. *Emerging Themes in Epidemiology*. 14:8 doi: 10.1186/s12982-017-0062-6

Raghunathan, T. (2016). *Missing Data Analysis in Practice*. Boca Raton: Taylor & Francis Group.

Reiter, J., Raghunathan, T., & Kinney, S. (2006). The Importance of Modeling Sampling Design in Multiple Imputation for Missing Data. *Survey Methodology, 32*(2), 143-149.

Rubin, Donald. (1987). *MI for Nonresponse in Surveys*. John Wiley & Sons, Inc.

Tshering, S., Okazaki, T., & Endo, S. (2013). A Method to Identify Missing Data Mechanism in Incomplete Dataset. *International Journal of Computer Science and Network Security, 13*(2). 14-22.

Van Buuren, S. Brands, J. P. L., & Groothuis-Oudshoorn, C. G. M. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*(12), 1049-1064.

Van Buuren, S. (2019). *Flexible Imputation of Missing Data*. Boca Raton: Taylor & Francis Group.

Zhu, J. & Raghunathan, T. (2015). Convergence Properties of a Sequential Regression Multiple Imputation Algorithm. *Journal of the American Statistical Association, 110*(511), 1112-1124.