RESEARCH REPORT SERIES
(Computing #2006-1)

The Predictive-Mean Method of Imputation for Preserving Coupling Between Assets and Liabilities

Yves Thibaudeau
Alfred Gottschalck[1]
Thomas Palumbo[1]

Statistical Research Division
U.S. Census Bureau
Washington, D.C.

Report Issued: October 23, 2006

[1]Housing and Household Economics Statistics Division

Executive Summary

In the context of Interagency Agreement (IAA) BC-04-03 between the Social Security Administration and the U.S. Bureau of the Census, the Census Bureau researched and developed an alternative to the present imputation strategy of using a univariate hot-deck methodology to impute assets and liabilities in the Wealth Topical Module for the Survey of Income and Program Participation (SIPP). Unlike the present strategy, the alternative imputation strategy (hereafter referred to as the "alternative strategy") explicitly accounts for the correlation between assets and liabilities at the household level by supplementing the univariate hot-deck with a joint hot-deck methodology and a model-based methodology referred to as *predictive-mean imputation.*

By blending of both statistical-match and model-based imputation methodology, the alternative strategy represents an intermediate step toward one of the goals of the IAA: developing a purely model-based approach to imputing assets and liabilities in SIPP. The development of the alternative strategy was guided by the requirement that it had to be able to be implemented within the existing SIPP processing system without a significant increase in processing costs or time.[1]

The primary difference between the present and alternative strategies is that the alternative uses the relationships between pairs of corresponding asset and liability variables (for example, *property value* and *mortgage amount)* to guide the imputation process. When a person reports the value of only one variable in such a pair, the alternative strategy uses this known information to impute a value for its missing companion (for example, in imputing values for the asset *property value*, the alternative strategy draws upon known information about the person's value for *mortgage amount*). When a person reports neither of the variables of a pair, the alternative strategy directly attempts to preserve the correlation between them by imputing them together from the same source (a donor in a joint hot-deck matrix).

---

[1] At this time, a new survey, the Dynamic of Economic Well-Being System (DEWS), is under design and will replace SIPP by 2009. DEWS, like SIPP, may carry a stand-alone wealth module. That is a module of questions on property, capital, equity and debts. So the timing of this report makes a logical starting point for the integrating of an alternative methodology along the lines of that described in this report. It is the intention of the authors to take full advantage of this opportunity.

We hypothesized: (1) that the alternative strategy would produce a substantially closer match than the present strategy between the correlations of the data for pairs of corresponding asset and liability variables when one or both of the items in the pair require imputation and the correlations when both items are reported; and (2) that a closer match would, in general, be associated with significant increases in the average imputed dollar amounts for the asset variables and significant decreases in those for the liability variables. We assumed that such movements in correlations and amounts would signify an improvement in the quality of the imputed wealth data and hence of the overall wealth data from SIPP, an assumption supported by the findings in the study by Mathematica Policy Research, Inc., that formed the basis for the Interagency Agreement.

For reasons of feasibility, we restricted our tests of these hypotheses to data for two pairs of corresponding asset and liability variables: (1) *property value/mortgage amount*; and (2) *business value/business debt*. These pairings represent opposite ends of the range of response rates for corresponding asset/liability pairings in SIPP: the two items in the first pair have relatively high rates of response; those of the second pairing relatively low response rates. They were selected because they likely typify the expected range of differences between the correlations and asset/liability amounts of the present and alternative strategies for SIPP asset/liability pairings in general.

The results of the tests suggest that the alternative strategy is an improvement over the present strategy. For both test pairings, the alternative strategy produced correlations for the imputed data that were substantially closer to those for the reported data than was true for the present strategy. The alternative strategy also generally raised asset amounts and lowered liability amounts. Above all, the findings demonstrated that taking account of the relationships between pairs of corresponding asset and liability variables in the imputation process could improve the quality of the SIPP wealth data. A potentially significant operational and statistical side benefit of the alternative strategy (particularly its predictive-mean component) is that it requires less manual intervention for its maintenance from one application to the next (wealth data are collected multiple times throughout a panel) than does the present strategy.

Given these favorable findings, we then examined the hypothesis that the alternative strategy is feasible to implement within the existing SIPP processing system. Our discussions with the processing staff suggested that the added costs and time associated with the new strategy may rule out its complete adoption until the 2009 SIPP panel. The need for the delay stems primarily from comparative deficiencies in the existing processing system owing to its dependence upon the obsolescent FORTRAN programming language; these should disappear when the processing is converted to the SAS© system for the 2009 panel.

In the interim, the processing staff thinks that it is feasible to introduce certain elements of the alternative strategy into the existing processing system to take advantage of the results of this research, beginning with the processing of the data for wave 3 of the 2004 panel. One revision would be to use a joint hot-deck method for the imputation of the

property value/mortgage amount and *business value/business debt* pairings when responses to both items in a pair are missing. Another is to account within the existing univariate hot-deck methodology for the asset/liability correlations of as many variable-pairings as is appropriate, by using the reported (or previously-imputed) value of one variable in a pair as a control (or "match") variable in the matrix that is used to impute the other. The research strongly suggested that substantial gains over the present imputation strategy could be expected from implementing these facets of the alternative strategy, if not the complete strategy.

Aware, then, of the likely benefits of the alternative strategy, but sensitive to the practical considerations against fully implementing it immediately, we recommend the following staggered approach to incorporating the results of this research into SIPP:

(1) To produce the official data from the 2004 panel, we recommend adapting the present imputation strategy as follows:

a) impute *property value/mortgage amount* and *business value/business debt* pairings jointly from the same source(donor) when both items in a pair are missing;
b) in the univariate hot-deck matrixes, use the reported value of one variable in the *property value/mortgage amount* and *business value/business debt* pair to impute the value of the other when only one is missing;
c) For as many other pairs of corresponding asset/liabilities variables as is appropriate, use either the reported or the imputed value of one member of the pair as a control variable in the unidimensional matrix that imputes the value for the other.

These adaptations will result in an "interim strategy" which represents a "partial" alternative strategy. This interim strategy preserves the joint imputation methodology of the alternative strategy, and its enhancements to the univariate matrixes, but discards its predictive-mean imputation methodology. As noted above, the research suggested that the preserved parts of the alternative strategy are capable, by themselves, of achieving many, if not all, of its comparative advantages over the present strategy.

(2) Beginning with the data for wave 3 of the 2004 panel, prepare research files produced by using the full alternative strategy (which includes predictive-mean imputation) for the *property value/mortgage amount* and *business value/business debt variables* (and perhaps other variable-pairs as well). The public-use versions of the records on these files should be able to be matched to those on the files for the corresponding official data. Production of these research files is to occur entirely outside the processing system for the official data from the 2004 panel. In this way, the research files surmount the timing and resource issues that argue against the application of predictive-mean imputation -- and hence of the full alternative strategy -- in the production of the official data.

(3) Use the above research files to evaluate whether the full alternative strategy (which includes predictive-mean imputation) is a significant improvement over the interim

strategy (which lacks predictive-mean imputation) presented in recommendation (1) above. This analysis can inform a decision on plans for improvements to the production of the wealth data in the 2009 Panel, including using the complete alternative strategy for other asset/liability pairs.

(4) If, using suitable criteria, the alternative strategy is judged in (3) above to represent a significant improvement, then it should be the exclusive one used for the wealth topical module for the 2009 SIPP panel, and should be applied to as many variable-pairs as appropriate. Research should be conducted on the desirability and feasibility of replacing the alternative strategy, which is a mixture of both model-based and statistical-match methodologies, with a purely model-based approach.

1. Introduction

The research for this project addressed the issues identified in the Mathematica Policy Research, Inc. report, "Survey Estimates of Wealth: A Comparative Analysis and Review of the Survey of Income and Program Participation,"[2] concerning the correlations between the data for corresponding pairs of asset and liability items in the wealth topical module of the Survey of Income Program Participation (SIPP). The levels of these correlations are key indicators of the quality of the imputed data. Improving them will serve the overall goal of the IAA to bring the estimates for the components of net worth in SIPP more in line with benchmark data from the Survey of Consumer Finances (SCF) of the Federal Reserve Board.

This report presents a detailed description of techniques for improving the quality of the imputed data. It analyzes and evaluates the results of applying these techniques to SIPP data, and provides recommendations. The research used data for month 4 of wave 6 of the 1996 SIPP Panel. The analysis focuses on the imputation of two asset-liability item pairs. The first pair consists of the variables *property value of primary residence (property value* for short) and its corresponding liability, *principal remaining on the mortgage (mortgage amount* for short). The second pair consists of *business value* and *debt against the business (business debt* for short). The two pairs were chosen because they represent the opposite ends of the spectrum of missing data situations encountered when processing SIPP wealth data.

The two items in the first pair have fairly high rates of response overall. The existing imputation process appears to *decouple* the relationship between them: that is, correlation between the two items is weaker when either or both items of the pair are imputed than when they are both reported. This inconsistency results from insufficient integration between the imputation processes for the items. Methods to remedy this problem are presented in this report.

Both items in the second pair have low response rates, and are more likely than those in the first pair to be both missing. Again, correlation between the items is weaker when either or both are imputed than when both are reported. Since they are often both missing,

[2] MPR Reference No. 8896-501

we propose a joint imputation procedure as a means of improving the quality of their imputations.

We illustrate our techniques for improving the imputation process by applying them to these two pairs of corresponding variables. We first identify and discuss the underlying causes of the observed correlation decoupling. Next, we describe the imputation strategies appropriate for countering their effects, and present the results of implementing these strategies. Lastly, we analyze and evaluate the results, and present our recommendations for improving the imputation system.

2. Background

In 2003, the Social Security Administration commissioned Mathematic Policy Research, Inc. (MPR), to conduct a comparative analysis and evaluation of differences between wealth estimates obtained from SIPP, with those from the Survey of Consumer Finances (SCF) of the Federal Reserve Board, and those from the Panel Study of Income Dynamics (PSID).

The MPR report pointed out the weak correlations in the 1996 SIPP panel between data for corresponding asset and liability variables (such as the asset variable property *value* and the liability variable *mortgage amount),* and it expressed concern over the degree to which the strength of these correlations has declined between the 1993 and 1996 SIPP panels. The report suggested that these phenomena may be indications that the imputation methodology – and the hot-deck imputation methods in particular – was a major source of the substantial differences between SIPP wealth estimates and those of the SCF and PSID.

The Census Bureau uses hot-deck imputation extensively when processing survey data. A clear advantage of the hot-deck method over other imputation methods (such as multiple imputation (Little and Rubin, 2002)) is the ease and simplicity of its implementation. Modern data analysis tools, however, have exposed shortcomings in this method, such as its potential to distort significantly the multivariate properties of data.

Evidence of such distortions in SIPP can be observed in wealth data for the 6^{th} wave of the 1996 panel. In particular, for the reported data, the correlation coefficient between property value and mortgage for is .39. But, it is only .16 (or less than half the fore-mentioned value), when only the property value is reported and the mortgage is imputed by the present hot deck. Across the entire sample, the correlation for the two items after imputation is .31, or 20 percent below the value of .39 for the jointly-reported data.

The apparent shortcomings of the hot-deck approach for the SIPP wealth data and their possible adverse affect on the quality of the data led to an interest in investigating whether a different method – particularly a model-based methodology such as that used in the SCF – could be developed that would produce better results and could be implemented to replace the hot-deck method. This interest was reflected in Goal 3 of IAA-BC-04-03, which stated that “the Census Bureau will research and develop a model-

based imputation approach for the imputation of assets and liabilities in the Wealth Topical Module [that] will account for the correlation between assets and liabilities at the household level… An intermediate step may be to account for this correlation by adapting the present procedure."

Pursuant to the intermediate step of this goal, we developed a strategy for modifying the existing imputation system, which we refer to as a *univariate hot-deck* strategy, by supplementing it with a *joint-hot-deck imputation* procedure, and a model-based method, known as *predictive mean imputation* (Little and Rubin 2002 p. 69). The predictive-mean component of the strategy may be considered to be a prototype for a more advanced model-based approach. The components of this strategy are described in detail in the next section, which also offers a critique of the present strategy.

This report assesses the results of a comparison between the outcomes of the present strategy with those of the alternative strategy. For the sake of operational feasibility, the study was limited to the two asset-liability pairings mentioned above, but owing to the range of imputation that these pairings represent, the results for them should be typical of what could be expected from applications of the process to other such pairings. The chosen variables are also of intrinsic interest and significance. Data for 2000 from SIPP found that home equity (the value of home property value minus mortgage amounts) constituted the largest share of household net worth, accounting for about a third of total net worth in 2000 (Orzechowski and Sepielli, 2003). A benchmark study comparing the 1998 SIPP estimate of business equity (business value minus business debt) with that of the 1998 SCF found a glaring difference between the two: a median in SIPP of $7,000 versus $60,000 for the SCF (Orzechowski and Sepielli, 2003).

3. Description of Imputation Methods

3.1 Univariate Hot-Deck Imputation

Univariate hot-deck imputation is the traditional imputation procedure employed by the Census Bureau. Only one variable at the time is imputed through this process. The procedure is based on an *imputation matrix*. The cells of the matrix are stacks. These stacks are *last-in, first-out* data structures where values of the variable to be imputed are stored. The rows and columns of the matrix are labeled by the value of the *class variables*. The class variables are categorical. The values stored in each cell are retrieved from records matching the values of the class variables for the specific row and column. Similarly, each imputation is retrieved from the cell matching the receiving record on the values of its class variables.

The univariate hot-deck process records in a geographical sort order and the cells of the matrix are filled in the same order. There can be any number of class variables. If there is only one, the matrix becomes a single row of cells. If there are more than two, the matrix becomes of multidimensional array of cells. For historical reason, we retain the term

"imputation matrix" regardless of the number of class variables and the dimensionality of the array.

For example, in the imputation of mortgage, reported values of mortgage are recorded and loaded into the imputation matrix. The loading sequence of donors into the cells is based on geography. Several cells of mortgage values are kept. Each cell corresponds to a different set of demographic characteristics, which define the identity of the donor and the recipient to a match. Whenever the process encounters a donor, it identifies the cell whose demographic characteristics correspond to those of this unit; it then places the value of that donor in that cell. Likewise, when the process encounters a recipient, it identifies the cell with corresponding demographic characteristics, and it assigns the mortgage value of the donor on top of this cell to the recipient.

Univariate hot-deck imputation can successfully handle many imputation problems. But some complex imputation problems, especially in a multivariate context, are poorly resolved by the method (Thibaudeau 2002). Limitations inherent in the univariate hot-deck procedure result in the following problems:

– *Shrunk Correlation:* Unless specifically designed, the univariate hot-deck reproduces correlations only indirectly. Two items are connected only through the characteristics corresponding to the cells in the imputation matrix. For example, the imputations of property value and mortgage could be processed individually, each based on a separate imputation matrix containing their respective values based on age, education, and occupation. If these demographic items are poor predictors of property value and/or mortgage, then the resulting imputed values likely will not reproduce the correlation between property value and mortgage for reported cases. Therefore, the correlation between property value and mortgage will lose its strength and will shrink: it gets close to 0.

– *Overly Narrow Imputation Cells:* In some cases, the imputation cells of a univariate hot-deck are narrowly defined (that is, defined by finely detailed breakdowns of a characteristic and/or by multiple characteristics). Such definition can lead to disproportionately narrow cells, meaning that a larger proportion of the people with missing values than of reported values fall into the cell. Narrow definition of imputation cells minimizes the bias and/or the random noise when substituting donor values for the missing items. The problem is, because the imputation cells are narrow, donors may not be available to fill all of them. To remedy this situation, default values, known as *cold deck* values, can be substituted. Cold-deck values have a tendency to be arbitrary. For example, the demographic characteristics defining the cells of an imputation matrix may be very specific. A cell could be defined by age (25 year old), education (PhD), and occupation (statistician). A unit with unreported mortgage with these demographic characteristics could be in the sample. But there may not be a unit with reported mortgage sharing these attributes. So a reported mortgage value is not available for substitution. Then a cold-deck (arbitrary) value is assigned.

– *Overly Broad Imputation Cells:* On the other hand, the imputation cells of the univariate hot-deck can be disproportionately large (that is, defined by broad breakdowns of a characteristic and/or by few characteristics) to maximize the chances of retrieving a donor. Then, the donor may substantially differ from the recipient, based on the stratification variables. For instance, the donor may be near the top of the range of values defining a cell, while the imputed unit may be at the bottom. An example is the imputation of mortgage. Because of the definition of the current imputation cells, donors with properties of considerable value are accepted in cells providing mortgage donors to sample units with modest property values. Consequently, some units subject to imputation can be attributed mortgages disproportionately large relative to the value of their property. If units with modest property values are over-represented among those missing a mortgage amount, the result is higher mortgages for these units.

3.2 Joint Hot-Deck Imputation

This method is designed to resolve some of the limitations of the univariate hot-deck imputation approach. Joint hot-deck imputation preserves the correlation between jointly missing items throughout the imputation process. For example, when an asset-liability pair, such as property value-mortgage, is missing, joint hot-deck imputation assigns to the receiver a property value and a mortgage from the same donor. In contrast, the univariate hot-deck methodology imputes the two items independently, often from separate donors. Consequently, joint hot-deck imputation is designed to maintain the multivariate properties between the two items, while the univariate hot-deck methodology is not.

3.3 Predictive Mean Imputation

3.3.1 An Intuitive Approach to the Predictive-Mean Methodology

Conceptually, the predictive mean method can be implemented by progressively expanding the number of cells of a one-dimensional hot-deck while maintaining cells of approximately equal sizes keeping at least one donor in each cell.

Table A illustrates this concept for the imputation of mortgage by using the reported value of property-value as the match variable. It shows the results obtained from our research by progressively increasing the number of cells of a unidimensional hot-deck. The definitions of the cells and their matrices in each case (one-cell, two-cell, four-cell and eight-cell matrices) are shown in Appendix A. The successive steps of the process are shown in columns 3 through 6. At each step, a new and more numerous set of cells is defined by further delineating subdivisions within the overall range of property values. The subdivisions are chosen so as to define cells of approximately equal size. For example, a two-cell stack might be defined by subdividing property values into two ranges: “$1 to $1000,” and “$1001 and over”; a four-cell stack by further subdividing the overall set of property values into four ranges: “$1 to $500,” $501 to $1000,” “$1001 to $1500,” and “$1501 and over;” and so on.

Table A. Illustration of Affinities between Univariate Hot-Deck and Predictive-Mean Imputation Methodologies

Measures	*Mortgage Reported* (col 2)	*Mortgage Imputed*				
		Univariate Hot Deck Method				*Predictive Mean Method* (col 7)
		One-cell Stack (col 3)	*Two-cell Stack* (col 4)	*Four-cell Stack* (col 5)	*Eight-cell Stack* (col 6)	
Approximate Cell Size (Number of Units to be Imputed in Each Cell)	N/A	1,156	578	289	144	1
Mean Mortgage Amount	$53,033	$53,394	$52,643	$50,550	$47,259	$46,957
Median Mortgage Amount	$35,000	$38,000	$38,000	$36,750	$30,000	$30,000
Coefficient of Correlation Between Property Value and Mortgage Amount	.39	.06	.14	.24	.26	.29

In the situation illustrated in the table, the univariate hot-deck with eight cells gives results close to those of the predictive mean approach. The mean and median property values, and the coefficient of correlation between property value and mortgage of the hot-deck, are all close to their predictive-mean counterparts.

Comparing column 6 and column 7 shows how a carefully constructed univariate hot-deck based on a unidimensional matrix can produce results similar to those of the predictive-mean method. Such results, however, can generally be achieved only if the cell sizes of the univariate hot-deck are relatively uniform, which means that new subdivisions of the values of the match variable must be defined each time the matrix is used for a new sample. This process requires repeated manual intervention in the imputation process. It is labor-intensive for both the analyst and the programmer, tedious, and as much an art as a science. A key advantage of the predictive mean method is that it eliminates the need for such interventions, since it automatically adjusts and optimizes the cell definitions.

3.3.2 Formal Development

The predictive mean method imputes mortgage without explicitly defining an imputation matrix. Rather, the method identifies single "best donors". To theoretically support this approach, several underlying assumptions are made and must be clarified.

In this section we develop a 3-stage model to provide a theoretical basis for using the predictive mean method. The first-stage of the model is overly naïve. The second-stage model is a refinement of the first. We show how this second model handles more general situations than the first. The third-stage model is a refinement of the second. It is the model underlying the fully functional predictive mean method.

A. Naïve Model

Let X and Y be two continuous variables. For the purpose of this application, X is mortgage and Y is property value. Suppose a sample of n units is taken yielding n observed (reported) values of Y denoted by $Y_1, Y_2, \ldots, Y_n$, where Y_i represents the value of Y for the ith household unit. Let the Y_i values be ordered as:

$$Y_1 < Y_2 < \ldots < Y_n \qquad \textbf{(1)}$$

Let Y_i represent the value of Y for the ith household unit. The values of Y are denoted as $Y_1, Y_2, \ldots, Y_n$. We assume the existence of a strictly increasing function $f(\square)$, for which the following holds

$$Y_i = f(X_i) \qquad \textbf{(2)}$$

In the context of (1) and (2), the predictive mean method imputation for X_i is $X_{\psi(i)}$ where $\psi(i)$ is the largest index such that $\psi(i) < i$ and $X_{\psi(i)}$ is reported. Then, X_i, which is the predictor of Y_i, is imputed by $X_{\psi(i)}$, which is the predictor of $Y_{\psi(i)}$.

Note, because $f(\square)$ is strictly increasing, $\psi(i)$ is always unique and $X_{\psi(i)} < X_i$.

We will see that only a few modifications to this definition of the predictive mean method are needed to cover more general situations.

B. More Realistic Model

Now, the relationship in (2) is unrealistic in the sense that there is no random component. A more realistic model is given next.

$$Y_i = f(X_i) + \varepsilon_i \qquad \textbf{(3)}$$

The ε_i's are uncorrelated random errors and have mean 0. We assume the Y_i's maintain their strict ordering $Y_1 < Y_2 < \ldots < Y_n$.

In the more realistic context of (3), the definition for the predictive mean we gave in the naïve context of (1) and (2) remains functional. But, while $\psi(i)$ remains unique, the proposition $X_{\psi(i)} < X_i$ no longer holds in general.

C. Most Realistic Model

To give a fully flexible representation of the imputation situations treated in this report, we introduce additional flexibility in our model. This final model accounts for a common situation: Round Dollar amounts, such as \$150,000, \$200,000, etc. are reported repeatedly. So, the assumption that the relationship between property value *Y* and mortgage *X* can be represented through a strictly increasing function $f(\square)$ is not.

We extend the model in (3) to obtain our most realistic and final model. For simplicity, we assume the repeated property value occur always at the same value, *C*. Our model is as follows.

$$Y_i = \begin{cases} C & \text{w. prob. } p \\ f(X_i) + \varepsilon_i & \text{w. prob. } 1-p \end{cases} \qquad \textbf{(5)}$$

For the sake of the presentation, let's assume *C* = \$150,000. The model in (5) can be interpreted as follows: The true value of some of these properties could be somewhat lower or higher than *C*. For example, suppose one property is worth \$147,000. Then, the respondent rounded-up the value of his property to \$150,000, perhaps because the information needed to make an accurate appraisal is not available to him.

We assume whenever a property value is in a neighborhood of C, the respondent report the value to be C "p of the time". The remaining "1-p of the time", the relationship between *X* and *Y* can be represented by a strictly increasing function $f(\square)$, as in the previous two models.

We extend the predictive mean method to deal with this situation:

1. When $Y_i = C$, the predictive mean method draws an imputation for X_i at random from the following set of reported property values: $\{X_j \mid Y_j = C\}$. This is equivalent to a random hot-deck. It will happen about "p of the time."
2. When $Y_i \neq C$ the predictive mean method remains as in stage 2. This will happen about "1-p of the time."

The predictive mean method will maintain coupling between X and Y. To achieve a comparable result, matrix-based hot-deck imputation requires narrow cells. When cells are disproportionately small, however, donors may not be available for many units. Complex collapsing schemes need to be implemented to recover units.

Because of its simplicity, the predictive mean method is easier to implement. It also guarantees results consistent with the model. This model is very general and covers many often encountered situations.

4. Alternative Imputation Strategy

The present strategy is to use the univariate hot-deck method to impute each of the asset and liability variables in a corresponding pair separately and without regard for the reported status (reported/not reported) of the value of the other variable.

Many other strategies are possible. We attempted to choose one that was model-based, though not necessarily exclusively, and that was feasible, in terms of costs and time, to implement within the existing SIPP processing system. The strategy we settled upon, and which we developed and evaluated in this research, chooses, for each of the variables in a pair, one of the above-three imputation methods, based upon the reporting status, and sometimes the level, of the value of its corresponding variable.

The following tables display the alternative strategy, for each of the two pairings we examined in our research:

Table B. Property Value (PV) and Mortgage (M)

Reporting Status	**Imputation Method Used**
Both PV and M missing	Joint Hot Deck
PV reported, M missing	Predictive Mean
PV missing, M reported, M=0	Univariate Hot Deck
PV missing, M reported, M>0	Predictive Mean

Table C. Business Value (BV) and Business Debt (BD)

Reporting Status	**Imputation Method Used**
Both BV and BD missing	Joint Hot Deck
BV missing, BD reported	Univariate Hot Deck
BV reported, BD missing	Univariate Hot Deck

5. Results of Applying Alternative Strategy

This section presents a comparative analysis and evaluation of the results of applying the present and alternative strategies to imputing the values for variables in the Property Value/Mortgage and Business Value/Debt pairings.

5.1 Property Value and Mortgage

Tables 1, 2, and 3 in Section 9 [3] show the results for the property value/mortgage asset-liability pair. Tables 1 and 2 show mean and median dollar amounts under both the present and alternative imputation strategies. Table 3 displays correlation coefficients.

In Table 1 under the present strategy (univariate hot-deck), median imputed property values are lower than the median for fully reported (values for both variables reported) cases, particularly so for the situation in which only the mortgage is reported and greater than zero ($91,500 versus $110,000). When both property value and mortgage are not reported, the imputed median property value is $8,000 less than that of fully reported cases ($102,000 versus $110,000).

Compared to the imputation of property value, the imputation data for mortgage amounts in Table 2 show the reverse relationship between the median dollar amounts for reported and imputed cases. Under the present strategy, median imputed values for cases where only property value is reported are $23,000 more than that of fully reported cases.

In addition, the correlation between property value and mortgage amount when only property value is reported is approximately half (.39 versus .16) that of fully reported cases (Table 3). When neither property value nor mortgage amount is reported, the correlation between the two items is 13 percent lower (.39 versus .34). As a result of the two patterns described above and the resulting lower correlation between property value and mortgage amount for cases imputed under the present methodology, property net worth is artificially depressed.

The alternative imputation strategy involves a mixture of the imputation methods described above (that is, univariate hot-deck, joint hot-deck, and predictive mean approaches). The alternative approach specifically incorporates the inherent correlation between property value and mortgage into the imputation process. Specifications for the imputation matrices employed are given in appendix B. The alternative imputation methodology is applied to four scenarios:

1. Both property value and mortgage are missing: joint hot-deck imputation of property value and mortgage is implemented (matrix A).
2. Property value is reported, but mortgage is missing: predictive mean imputation of mortgage is implemented based on property value.
3. Property value is missing, mortgage is reported, and mortgage amount is equal to 0: univariate hot-deck imputation of property value is implemented (matrix B).
4. Property value is missing, mortgage is reported, and mortgage amount is greater than 0: predictive mean imputation of property value is implemented based on mortgage.

The results from applying the alternative imputation strategy are shown in the bottom panels of Tables 1 and 2 and the bottom row of Table 3:

[3] For reasons of disclosure avoidance, only derived measures are shown in the tables in Section 9.

For scenario 1, the median value of property increased and mortgage amount decreased to a more accurate representation of the values for fully reported cases: $110,000 under the new strategy versus $102,000 under the present strategy for property value, and $44,000 versus $58,709 for mortgage amount. Furthermore, the correlation between the two items improved from .34 to .50, a 47 percent increase. Scenario 1 comprises 42 percent of the cases eligible for imputation.

For scenario 2 (23 percent of imputation eligible cases), the median mortgage amount decreased almost 50 percent, from $58,000 to $30,000. The correlation between property value and mortgage amount improved from .16 to .29, an 80 percent improvement. Under the present imputation strategy the correlation between the asset-liability pair was less than half that as for fully reported cases, while under the alternative strategy (predictive mean imputation) it was 74 percent of that for fully reported cases.

For scenario 3 (31 percent of imputation eligible cases), the alternative imputation strategy increases the median property value from $82,000 to $90,000, a 10 percent increase.

Lastly, the dollar figure results for scenario 4 (4 percent of imputation eligible cases) mirror those from scenario 3, where median property values increased, in this case by 17 percent ($91,500 to $107,500). The correlation between property value and mortgage amount increased slightly from .28 to .30, a 7 percent increase.

Over the entire distribution (both reported and imputed cases), the alternative imputation strategy increases the correlation between property value and mortgage by 19 percent (.37 versus .31), thereby more closely resembling the correlation (.39) of fully reported cases (Table 3).

5.2 Business Value and Business Debt

Tables 4 through 6 (section 9) show the results for the business value/business debt pairing for *all* businesses. Tables 4 and 5 show mean and median dollar amounts under both the present and alternative imputation strategies. Table 6 displays correlation coefficients. Tables 7, 8, and 9 present corresponding results for businesses with more than 25 employees,[4] but particular caution must be exercised in interpreting them because some figures are based on a relatively small number of observations.

In Table 4, under the present strategy, median imputed business values are lower than the median value for fully reported cases. When only business debt is reported, the median imputed business value is 60 percent lower than that of fully reported cases, and when both business value and debt are unreported, the median imputed business value is 50

[4] Results are not shown separately for businesses of 25 or fewer employees. Such businesses constitute 96 percent of the observations for businesses in this study (1,955 out of 2,028). Hence, results for them should be virtually identical to those for all businesses, and their presentation would add little of value to the discussion.

percent lower than that of fully reported cases. The second situation comprises 74 percent of the imputation eligible cases. For business debt, Table 5 shows that under the present imputation strategy higher debt values are imputed when both business value and debt are missing than when both are fully reported.

For businesses with more than 25 employees, Table 7 shows that the median imputed business value when both business value and debt are unreported [5] is only about 10 percent of the median for fully reported cases ($18,000 versus $170,000). Table 8 displays the corresponding imputation outcomes for business debt. When both business value and business debt are unreported, the median imputed business debt value is higher ($5,000 versus $0) than when both items are reported.

As for the correlations between business value and debt for all businesses and for businesses with more than 25 employees (Tables 6 and 9, respectively), under the present strategy correlations are anywhere between 50 percent and 90 percent lower compared to the correlations observed for fully reported cases. As was the case for the property value/mortgage pairing, the resulting lower correlation between business value and debt for imputed cases and the lower imputed median dollar value for business value artificially depress business net worth.

The alternative imputation strategy differs from the present strategy in two major ways. First, the alternative strategy imputes business value and debt jointly when they are both missing, which covers the majority (74 percent) of cases where imputation is needed. Second, the alternative strategy uses two sets of imputation matrices. The first set of imputation matrices is designed for imputing business value and debt for businesses with up to 25 employees. The second set is designed for businesses with more than 25 employees. To facilitate the retrieval of donors and to minimize cold-deck imputation, the matrices in the second set have larger cells than those of the first set. The specifications for these imputation matrices are presented in Appendix C.

The alternative imputation strategy is applied to the following three scenarios:

1. Business value and debt are both missing: joint hot-deck imputation is implemented. If the number of employees is 25 or fewer, matrix C is used; otherwise, matrix D is used.
2. Business Value is reported and debt is not reported: univariate hot-deck imputation is used. If the number of employees is 25 or fewer, matrix E is used; otherwise matrix F is used.
3. Business value is missing and debt is reported: univariate hot-deck imputation is used. If the number of employees is 25 or fewer, matrix G is used; otherwise, matrix H is used.

The results from applying the alternative imputation strategy are shown in the bottom panels of Tables 4, 5, 7, and 8 and the bottom row of Tables 6 and 9:

[5] Approximately 90 percent of imputation-eligible cases fall into this category.

For scenario 1, both categories of businesses (all businesses as a whole and businesses with more than 25 employees) saw increases in their median business value figures (Tables 4 and 7): $5,000 to $15,000 for businesses as a whole and $18,000 to $140,000 for businesses with more than 25 employees. Both medians are representative of the reported population. For scenario 2, median business value for businesses as a whole increased 200 percent ($4,000 to $12,000), while for businesses with more than 25 employees the median figure decreased 66 percent ($235,000 to $80,000).

As for business debt (Tables 5 and 8), for businesses as a whole the alternative strategy produced lower debt figures in scenario 1 that are more comparable to the fully reported cases. For businesses with more than 25 employees, the alternative strategy produced a higher median debt figure compared with the present imputation methodology. The results for scenario 3 show little difference in median values between the present and alternative imputation strategies for both business groups.[6]

For the distribution as a whole (all cases), the alternative imputation strategy more than doubles the correlation between business value and debt for both business groups (Tables 6 and 9). The correlations for both business groups are now more representative of the correlations observed for fully reported cases.

5.3 Marginal Effects on Overall Estimates of Value of Property and Business Net Worth

Table 10 compares the impacts of the present and alternative imputation strategies on the contribution of property and businesses to a household's net worth. The figures for *marginal net worth of property* represent property value minus mortgage amount; and those for *marginal net worth of business* equal business value minus business debt. The term *marginal* signifies that the figures represent additions to overall household net worth.

The table shows that the alternative imputation strategy produces higher median and mean property and business net worth values than the present strategy, particularly for businesses with more than 25 employees. The median and mean property net worth values increased 3 percent and 8 percent, respectively. For businesses as a whole, median business net worth doubled and the mean business net worth value increased 2 percent. As for businesses with more than 25 employees, the alternative strategy produced median and mean business net worth values significantly greater than those produced under the current strategy: $140,000 versus $15,000 and $874,779 versus $331,321, respectively.

5.4. Summary of Results

The following two sections summarize the results for each asset-liability pair.

5.4.1 Imputation of Property Value/Mortgage

[6] Note that the number of observations for this scenario is very small.

There are two salient differences between the results obtained from the present and alternative approaches:

1. The alternative imputation strategy substantially increases the overall correlation between property value and mortgage. The new correlation is similar to that observed when examining only fully reported cases.
2. The alternative strategy produces a substantially lower mean mortgage relative to the present strategy. This has an impact on the marginal net worth associated with the property. Distributional descriptors of the marginal net worth associated with property are given in Table 10. The median and mean marginal net worth figures are higher under the alternative imputation strategy.

5.4.2 Imputation of Value of Business/Debt

1. The alternative strategy produces a correlation substantially closer to the reported correlation than that obtained using the present strategy.
2. For all businesses, the median and mean marginal net worth figures are higher under the alternative imputation strategy than under the present strategy; for businesses with more than 25 employees, the corresponding figures for the alternative strategy are substantially higher than those for the present strategy (Table 10).

6. Implementation

An important consideration in designing the alternative strategy was ease of implementation within the existing SIPP processing system. We designed the alternative strategy to possess this property. Its univariate hot-deck elements already exist in the present strategy. The use of joint hot-decks to accompany the univariate hot-decks is well established in SIPP. Joint hot-decks do not involve any more processing time than univariate hot-decks; the cost of programming them should be minimal, since they can usually be adapted readily from their already-existing univariate counterparts. As we have developed it, the only potential drawback of the predictive-mean methodology is that it requires a resorting of the entire file each time a new variable is to be imputed. But file sorting is an easy and fast procedure to implement within the SAS programming system that will be used to carry out the processing, beginning with the 2009 panel. An especially favorable factor is that the alternative strategy is intended for use for a topical module, rather than for a section of the core elements of SIPP, so any additional processing time required by the alternative strategy should be a secondary and non-deciding consideration. The alternative's potential reduction in manual intervention and maintenance costs over repeated applications is a particularly attractive benefit.

7. Conclusion

The report from MPR suggests that the wealth estimates based on SIPP data are distorted by the present imputation process. The distortions likely stem from holes in the design of the current imputation methodology employed to process SIPP wealth data. The most important limitation of the current imputation methodology is the absence of features for reproducing the multivariate properties (correlation) between specific assets and liabilities.

This report analyzed the multivariate aspect of two pairs of wealth items from the 1996 panel of SIPP. Based on our analysis, we developed and tested an alternative imputation methodology with the goal of preserving the multivariate properties of the data throughout the imputation process. The alternative represents a modification of the present imputation system. Basically, it supplements the present system with a joint hot-deck-imputation procedure and a simple model-based methodology known as predictive mean imputation. We took special care to design adaptable methods that can be realistically implemented within the context of the current SIPP processing system.

The results from our evaluation show that the alternative methodology can significantly improve the imputed data for the two asset-liability pairs we examined. The evidence for this improvement is that, under the alternative methodology, the means, medians, and correlations of the imputed data are more representative of the fully reported data than they are under the present strategy.[7] As a consequence, the overall property and business net worth values of the alternative are higher than those of the present methodology.

8. Recommendations

(1) To produce the official data from the 2004 panel, we recommend adapting the imputation present strategy as follows:

1. impute *property value/mortgage amount* and *business value/business debt* pairings jointly from the same source(donor) when both items in a pair are missing;
2. in the univariate hot-deck matrixes, use the reported value of one variable in the *property value/mortgage amount* and *business value/business debt* pair to impute the value of the other when only one is missing;
3. for as many other pairs of corresponding asset/liabilities variables as is appropriate, use either the reported or the imputed value of one member of the pair as a control variable in the univariate matrix that imputes the value for the other.

[7] This statement assumes that the distributions of nonrespondents (cases needing imputation) by the relevant variables are similar to those of respondents (cases reporting data). It also assumes that, in general, the greater the conformity between the imputed data and the fully-reported data, the better the quality of the imputed data and of the overall data. We treated these assumptions as axioms of our research.

(2) Beginning with the data for wave 3 of the 2004 panel, prepare research files containing data produced by using the alternative strategy, including the predictive mean methodology, for the *property value/mortgage amount* and *business value/business debt variables* (and perhaps other variable-pairs as well); the public-use versions of the records on these files should be able to be matched to those on the files for the corresponding official data.

(3) Use the above research files to evaluate whether the alternative strategy is a significant improvement over the interim approach described in (1) above. This analysis can inform a decision on plans for improvements to the production of the wealth data in the 2009 Panel, including using the complete alternative strategy for other asset/liability pairs.

(4) If, using suitable criteria, the alternative strategy is judged in (3) above to represent a significant improvement, then it should be the exclusive one used for the wealth topical module for the 2009 SIPP panel, and should be applied to as many variable-pairs as appropriate. Research should be conducted on the desirability and feasibility of replacing the alternative strategy, which is a mixture of both model-based and statistical-match methodologies, with a purely model-based approach.

9. Data Tables

Table 1 – Summary Property Value Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Property Value (PV) and Mortgage (M) (dollars)

Type of Strategy and Summary Measurements	*Both PV and M Reported*	*Only PV Reported*	*Only M Reported And M = 0* *(Scenario 3)*	*Only M Reported and M > 0* *(Scenario 4)*	*Neither PV nor M Reported* (Scenario 1)	*All Cases*
Imputation Method						
Present Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck			Not Applicable
Mean	138,206	138,543	106,306	124,761	135,271	134,225
Median	110,000	100,000	82,000	91,500	102,000	100,000
Top Decile	250,000	250,000	180,000	225,000	250,000	250,000
Alternative Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck	Predictive Mean	Joint Hot-deck	Not Applicable
Mean	138,206	138,543	115,177	140,898	136,906	135,642
Median	110,000	100,000	90,000	107,500	110,000	107,000
Top Decile	250,000	250,000	200,000	240,000	250,000	250,000

Table 2 – Summary Mortgage-Value Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Property Value (PV) and Mortgage (M) (dollars)

Type of Strategy and Summary Measurements	*Both M and PV Reported*	*Only PV Reported* *(Scenario 2)*	*Only M Reported*	*Neither PV nor M Reported* (Scenario 1)	*All Cases*
Imputation Method					
Present Strategy	Not Applicable	Univariate Hot-Deck	Not Applicable	Univariate Hot-Deck	Not Applicable
Mean	53,033	77,115	7,054	73,636	52,371
Median	35,000	58,000	0	58,709	35,000
Top Decile	133,000	134,000	60,000	142,000	128,046
Alternative Strategy	Not Applicable	Predictive Mean	Not Applicable	Joint Hot-Deck	Not Applicable
Mean	53,033	46,957	7,054	57,090	47,569
Median	35,000	30,000	0	44,000	25,000
Top Decile	133,000	120,000	60,000	130,000	125,000

Table 3 – Correlation Coefficients between Property Value and Mortgage, by Present and Alternative Imputation Strategies

Imputation Strategies	*Both PV and M Reported*	*Only PV Reported* *(Scenario 2)*	*Only M Reported And MV = 0* *(Scenario 3)*	*Only M Reported and M > 0* *(Scenario 4)*	*Neither PV nor M Reported* (Scenario 1)	*All Cases*
Present Strategy	.39	.16	Not Applicable	.28	.34	.31
Alternative Strategy	.39	.29	Not Applicable	.30	.50	.37
Percentage of Eligibles for Imputation	Not Applicable	23 %	31 %	4 %	42 %	NA

Table 4 – Summary Value of Business Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Business Value (BV) and Debt (D) for All Businesses (dollars)

Type of Strategy and Summary Measurements	*Both BV and D Reported*	*Only BV Reported*	*Only D Reported* *(Scenario 2)*	*Both BV and D Unreported* *(Scenario 1)*	*All Cases*
Imputation Method					
Present Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck		Not Applicable
Mean	113,590	428,948	140,337	154,153	137,325
Median	10,000	60,000	4,000	5,000	10,000
Top Decile	300,000	750,000	300,000	250,000	300,000
Alternative Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck	Joint Hot-Deck	Not Applicable
Mean	113,590	428,948	128,658	160,554	138,277
Median	10,000	60,000	12,000	15,000	12,000
Top Decile	300,000	750,000	250,000	500,000	353,000

Table 5 – Summary Debt Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Business Value (BV) and Debt (D) for All Businesses (dollars)

Type of Strategy and Summary Measurements	*Both BV and D Reported*	*Only BV Reported* *(Scenario 3)*	*Only D Reported*	*Both BV and D Unreported* *(Scenario 1)*	*All Cases*
			Imputation Method		
Present Strategy	Not Applicable	Univariate Hot-Deck	Not Applicable	Univariate Hot/Deck	Not Applicable
Mean	20,753	16,996	9,705	30,890	23,295
Median	0	0	0	3,000	0
Top Decile	45,000	100,000	10,000	92,000	60,000
Alternative Strategy	Not Applicable	Univariate Hot-Deck	Not Applicable	Joint Hot-Deck	Not Applicable
Mean	20,753	43,797	9,705	25,996	21,696
Median	0	10,000	0	0	0
Top Decile	45,000	138,000	10,000	50,000	45,000

Table 6 –Correlation Coefficients between the Value of Business (BV) and Debt (D) for All Businesses

Imputation Strategies	*Both BV and D Reported*	*Only BV Reported* *(Scenario 3)*	*Only D Reported* *(Scenario 2)*	*Both BV and D Unreported* *(Scenario 1)*	*All Cases*
Present Strategy	.46	.13	0	.15	.21
Alternative Strategy	.46	0	.26	.59	.48

Table 7 – Summary Value of Business Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Business Value (BV) and Debt (D) for Businesses with More than 25 Employees (dollars)

Type of Strategy and Summary Measurements	*Both BV and D Reported*	*Only BV Reported*	*Only D Reported (Scenario 2)*	*Both BV and D Unreported (Scenario 1)*	*All Cases*
Imputation Method					
Present Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck		Not Applicable
Mean	892,500	5,000,000	407,500	174,314	377,643
Median	170,000	5,000,000	235,000	18,000	40,000
Top Decile	2,000,000	5,000,000	1,000,000	265,000	950,000
Alternative Strategy	Not Applicable	Not Applicable	Univariate Hot-Deck	Joint Hot-Deck	Not Applicable
Mean	892,500	5,000,000	736,666	925,918	953,235
Median	170,000	5,000,000	80,000	140,000	140,000
Top Decile	2,000,000	5,000,000	4,000,000	4,000,000	4,000,000

Table 8 – Summary Debt Measurements, by Present and Alternative Imputation Strategies, by Reported-Status of Business Value (BV) and Debt (D) for Businesses with More than 25 Employees (dollars)

Type of Strategy and Summary Measurements	*Both BV and D Reported*	*Only BV Reported* *(Scenario 3)*	*Only D Reported*	*Both BV and D Unreported* *(Scenario 1)*	*All Cases*
Imputation Method					
Present Strategy	Not Applicable	Univariate Hot-Deck	Not Applicable	Univariate Hot/Deck	Not Applicable
Mean	83,750	0	64,166	36,880	46,322
Median	0	0	0	5,000	2,000
Top Decile	200,000	0	325,000	104,000	120,000
Alternative Strategy	Not Applicable	Univariate Hot-Deck	Not Applicable	Joint Hot-Deck	Not Applicable
Mean	83,750	0	64,166	94,387	88,455
Median	0	0	0	25,000	0
Top Decile	200,000	0	325,000	750,000	325,000

Table 9 – Correlation Coefficients between the Value of Business (BV) and Debt (D) for Businesses with More than 25 Employees

Imputation Strategies	*Both BV and D Reported*	*Only BV Reported* *(Scenario 3)*	*Only D Reported* *(Scenario 2)*	*Both BV and D Unreported* *(Scenario 1)*	*All Cases*
Present Strategy	.84	Not Applicable	.41	.057	.31
Alternative Strategy	.84	Not Applicable	.98	.86	.77

Table 10 – Comparison of Marginal Net-Worth Measures (dollars) of Property and of Businesses, by Present and Alternative Imputation Strategies

Strategies	*Mean*	*Bottom 1%*	*Bottom 10%*	*Median*	*Top 10%*	*Top 1%*
Marginal Net-Worth of Property						
Present Strategy (a)	81,853	-32,300	10,000	60,000	174,000	487,000
Alternative Strategy (b)	88,056	0	14,000	62,000	180,000	500,000
Percent Change	7.6	Above 100	40.0	3.3	3.5	2.7
Marginal Net-Worth of All Businesses						
Present Strategy (a)	114,030	-149,000	-8,000	5,000	250,000	2,000,000
Alternative Strategy (b)	116,580	-84,000	0	10,000	260,000	2,000,000
Percent Change	2.2	43.6	Above 100	100	4.0	No change
Marginal Net-Worth of Businesses With More than 25 Employees						
Present Strategy (a)	331,321	-600,000	-31,960	15,000	950,000	5,000,000
Alternative Strategy (b)	874,779	0	0	140,000	3,250,000	5,000,000
Percent Change	Above 100	Above 100	Above 100	Above 100	Above 100	No change

Notes:

(1) "Bottom 1%" is the value of the upper limit of the 1% percentile; "Bottom 10%" is the value of the upper limit of the 10% percentile; "Top 10%" is the value of the lower limit of the 90% percentile; "Top 1%" is the value of the lower limit of the 99% percentile.

(2) "Percent change" equals ((b-a)/a) x 100.

References

Little, J. A. and D.B. Rubin (2002). "Statistical Analysis with Missing Data." Wiley.

Mathematica Policy Research, Inc. (2003). "Survey Estimates of Wealth: A Comparative Analysis and Review of the Survey of Income and Program Participation."

Orzechowski, Shawna and Peter Sepielli (2003). " Net Worth and Asset Ownership of Households: 1998 and 2000," U.S. Census Bureau, Current Population Reports, P70-88.

Thibaudeau, Y. T. (2002). "Model Explicit Item Imputation for Demographic Categories," Survey Methodology, *28*, 2.

Appendix A. Imputation Matrices for the Predictive Mean Example

1 Cell:

This imputation matrix has one imputation cell for any reported property value.

2 Cells:

Variable	*Values (dollars)*
Reported Property Value	1) 0 – 99,999 2) 100,000 or more

4 Cells:

Variable	*Values (dollars)*
Reported Property Value	1) 0 – 69,999 2) 70,000 – 99999 3) 100,000 – 159,999 4) 160,000 or more

8 Cells:

Variable	*Value (dollars)*
Reported Property Value	1) 0 – 49,999 2) 50,000– 69,999 3) 70,000 – 88,999 4) 89,000 – 99,999 5) 100,000 – 134,999 6) 135,000 – 159,999 7) 160,000 – 204,999 8) 205,000 or more

Appendix B. Imputation Matrices – Property Value and Mortgage Amount

Matrix A

Joint imputation of property value and mortgage amount:

Variable	*Values*
Race	1) Black 2) Nonblack
Age (years)	1) 0 – 29 2) 30 – 44 3) 45 – 64 4) 65 or above
Quarterly Income (dollars)	1) 0 – 2,999 2) 3,000 – 4,999 3) 5,000 – 9,999 4) 10,000 – 18,499 5) 18,500 or more

Matrix B

Imputation of property value when mortgage amount is zero:

Variable	*Values*
Race	1) Black 2) Nonblack
Age (years)	1) 0 – 29 2) 30 – 44 3) 45 – 64 4) 65 or above
Quarterly Income (dollars)	1) 0 – 2,999 2) 3,000 – 4,999 3) 5,000 – 9,999 4) 10,000 – 18,499 5) 18,500 or more

Appendix C. Imputation Matrices – Business Value and Business Debt

Matrix C

Joint imputation of business value and debt – 25 or fewer employees:

Variable	*Values*
Property Value of Primary Residence (dollars)	1) 0 – 149,999 2) 150,000

Matrix D

Joint imputation of business value and debt – more than 25 employees:

This imputation matrix has one cell: donors and recipients will be matched solely on the basis of the sort order of the processing of the records

Matrix E

Imputation of debt – 25 or fewer employees

Variable	*Values*
Business Value (dollars)	1) 0 – 499,999 2) 500,000 or more

Matrix F

Imputation of debt – more than 25 employees:

This imputation matrix has one cell: donors and recipients will be matched solely on the basis of the sort order of the processing of the records

Matrix G

Imputation of business value – 25 or fewer employees:

Variable	*Values*
Debt Held Against the Business (dollars)	1) 0 – 299,000 2) 300,000 or more

Matrix H

Imputation of business value – more than 25 employees:

This imputation matrix has one cell: donors and recipients will be matched solely on the basis of the sort order of the processing of the records