Fractional hot deck imputation for robust inference under item nonresponse in survey sampling

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Shu Yang
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_pubs
Part of the Design of Experiments and Sample Surveys Commons, and the Statistical Methodology Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/stat_las_pubs/116. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.
Fractional hot deck imputation for robust inference under item nonresponse in survey sampling

Jae Kwang Kim and Shu Yang

Abstract

Key Words: EM algorithm; Kullback-Leibler information; Missing at random (MAR); Multiple imputation.

1 Introduction

Imputation is a popular method of compensating for item non-response in sample surveys. Let $y$ be the study variable subject to non-response and $x$ be the vector of auxiliary variables fully observed. A model on the conditional distribution $f(y|x)$ is often used to generate imputed values for missing $y_i$. Such model-based imputation method is well developed in the literature. Multiple imputation of Rubin (1987) is a Bayesian approach of model-based imputation. Monte Carlo EM of Wei and Tanner (1990) can be treated as a frequentist's approach of model-based imputation. Kim (2011) proposed parametric fractional imputation to handle multivariate missing data.

However, the model-based imputation method that generates imputed values from $f(y|x)$ is not a hot deck imputation in the sense that artificial values are constructed after the imputation. A desirable property of hot deck imputation is that all imputed values are observed values. For example, imputed values for categorical variables will also be categorical with the same number of categories as observed for the respondents. For this reason, hot deck imputation is the most popular imputation method, especially in household surveys. Nearest neighbor imputation method is also a hot deck imputation. Chen and Shao (2001), Beaumont and Bocci (2009), Kim, Fuller and Bell (2011) investigated nearest neighbor imputation in the context of survey sampling. Durrant (2009), Haziza (2009) and Andridge and Little (2010) provided comprehensive overviews of the hot-deck imputation methods in survey sampling.

Fractional hot deck imputation was proposed by Kalton and Kish (1984) to achieve efficiency in hot deck imputation. Kim and Fuller (2004) and Fuller and Kim (2005) provided a rigorous treatment of fractional hot deck imputation and discussed variance estimation. However, their approach is only applicable when $x$ is categorical. For continuous covariate case, predictive mean matching can be treated as a nearest neighbor imputation method using the predicted value obtained from $f(y|x)$ but its statistical properties are not fully addressed in the literature.

1. Jae Kwang Kim, Department of Statistics, Iowa State University, Ames, IA 50011. E-mail: jkim@iastate.edu; Shu Yang, Department of Statistics, Iowa State University, Ames, IA 50011.
In this paper, we propose a new fractional hot deck imputation (FWDI) method based on a parametric model of $f(y|\mathbf{x})$ that allows continuous covariates. The proposed method has several advantages over the existing methods. First, it is a hot deck imputation preserving the correlation structure between the items. Second, it is robust in that the resulting estimator is less sensitive against the failure of the assumed model $f(y|\mathbf{x})$. Third, it provides consistent variance estimators for various parameters without requiring the congeniality condition of Meng (1994). Multiple imputation, however, requires the congeniality condition for the validity of variance estimation. When the congeniality condition does not hold, multiple imputation often leads to conservative inference, which in turn reduces test powers. See Section 5.2 for more details.

The paper is organized as follows. Section 2 describes the basic setup. The proposed method is presented in Section 3. The robustness of FWDI is discussed in Section 4. Results from two simulation studies are presented in Section 5 before some concluding remarks are made in Section 6.

2 Basic setup

Consider a finite population of $N$ elements identified by a set of indices $U = \{1, 2, \ldots, N\}$ with $N$ known. Associated with each unit $i$ in the population are study variables, $x_i$ and $y_i$, with $x_i$ always observed and $y_i$ subject to non-response. Let $A$ denote the set of indices for the elements in a sample selected by a probability sampling mechanism. We are interested in estimating $\eta$, defined as a (unique) solution to the population estimating equation

$$\sum_{i=1}^{N} U(\eta; x_i, y_i) = 0.$$  

(2.1)

where $w_i = \left\{Pr(i \in A)\right\}^{-1}$ is the inverse of the first-order inclusion probability of unit $i$. Binder and Patak (1994) and Rao, Yung and Hidiroglou (2002) considered the asymptotic properties of the estimator obtained from (2.1). Under the existence of missing data, we define

$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$

A consistent estimator of $\eta$ is then obtained by taking the conditional expectation and solving

$$\sum_{i \in A} w_i \left[\delta_i U(\eta; x_i, y_i) + (1 - \delta_i)E\{U(\eta; x_i, Y)|x_i, \delta_i = 0\}\right] = 0$$  

(2.2)

for $\eta$. Estimating equation (2.2) is sometimes referred to as expected estimating equation (Wang and Pepe 2000).

To compute the conditional expectation in (2.2), we assume that the finite population at hand is a realization from an infinite population, called superpopulation. In the superpopulation model, we often postulate a parametric conditional distribution of $y$ given $x$, $f(y|x; \theta)$, which is known up to the parameter $\theta$ with parameter space $\Omega$. Under the specified model, we can compute a consistent estimator
\( \hat{\theta} \) of \( \theta \) and then use a Monte Carlo method to evaluate the conditional expectation in (2.2) given the estimate \( \hat{\theta} \). If the response mechanism is missing at random (MAR) or ignorable in the sense of Rubin (1976), we can approximate the expected estimating equation in (2.2) by

\[
\sum_{i \in A} w_i \left\{ \delta U(\eta; x_i, y_i) + (1 - \delta_i) \frac{1}{m} \sum_{j=1}^m U(\eta; x_i, y_i^{(j)}) \right\} = 0,
\]

(2.3)

where

\[
y_i^{(1)}, \ldots, y_i^{(m)} \sim f(y_i | x_i; \hat{\theta}).
\]

Often, we use the maximum likelihood estimator \( \hat{\theta} \), which solves

\[
S(\theta) = \sum_{i \in A} w_i \delta_i S(\theta; x_i, y_i) = 0,
\]

(2.4)

where \( S(\theta; x, y) = \partial \log f(y | x; \theta) / \partial \theta \). Note that we use the sampling weights \( w_i \) in the score equation (2.4). Thus, we are implicitly assuming that the imputation model, the model for generating the imputed values, is the model about the finite population values \( f(y_i | x_i) \), not the model about the sample values. Thus, we allow that the sampling mechanism can be informative in the sense of Pfeffermann (2011). Multiple imputation, on the other hand, uses the sample model, \( f_i(y_i | x_i) \equiv f(y_i | x_i, i \in A) \), to generate the imputed values and often assumes that the sampling mechanism is non-informative. Thus, in multiple imputation, MAR is assumed for the sample at hand, while, in fractional imputation, MAR is assumed for the population. Under informative sampling design, generating imputed values from the sample model \( f_i(y_i | x) \) does not necessarily lead to valid inference even when sample MAR condition holds. See Section 8.4 of Kim and Shao (2013) for further discussion of MAR under informative sampling.

To compute the conditional expectation in (2.2) efficiently, the parametric fractional imputation (PFI) of Kim (2011) can be used. In PFI, the imputed values are generated from a suitable proposal distribution \( h(y | x_i) \) and then the imputed estimating equation (2.3) is changed to

\[
\sum_{i \in A} w_i \left\{ \delta U(\eta; x_i, y_i) + (1 - \delta_i) \sum_{j=1}^m w_j^* U(\eta; x_i, y_i^{(j)}) \right\} = 0,
\]

(2.5)

where

\[
w_j^* = \frac{f(y_i^{(j)} | x_i; \hat{\theta})/h(y_i^{(j)} | x_i)}{\sum_{k=1}^m \left\{ f(y_i^{(k)} | x_i; \hat{\theta})/h(y_i^{(k)} | x_i) \right\}}.
\]

(2.6)

The choice of the proposal distribution \( h(\cdot) \) is somewhat arbitrary. We will discuss a particular choice that may lead to a robust estimation.

The consistency of the resulting estimator \( \hat{\eta} \) from (2.3) or (2.5) can be established under the assumption that the conditional distribution \( f(y | x; \theta) \) is correctly specified (by similar argument in the proof of Corollary II.2 of Andersen and Gill (1982) and its proof is skipped here). In this paper, we consider an alternative approach of fractional imputation that is more robust against the failure of the assumption on the imputation model.
3 Proposed method

We first consider a particular fractional hot deck imputation method, called **full fractional imputation**, where the imputed values are taken from the set of respondents denoted as \( A_r = \{ i \in A; \delta_i = 1 \} \). That is, the \( j \)-th imputed value of missing \( y_i \), denoted by \( y_i^{(j)} \), is equal to the \( j \)-th value of \( y \) among the set in \( A_r \). We propose a fractional hot deck imputation approach that makes use of the parametric model assumption \( f(y|x; \theta) \). If all of the elements in \( A_r \) are selected as the imputed values for missing \( y_i \), we can treat \( \{ y_j; j \in A_r \} \) as a realization from \( f(y_j|\delta_j = 1) \) and fractional weight assigned to donor \( y_j \) for the missing item \( y_i \) is, by choosing \( h(y_j|x) = f(y_j|\delta_j = 1) \) in (2.6),

\[
\begin{align*}
  w^*_{iy} &\propto \frac{f(y_j|x, \delta_j = 0; \hat{\theta})}{f(y_j|\delta_j = 1)} \\
  &\propto \frac{f(y_j|x, \hat{\theta})}{f(y_j|\delta_j = 1)},
\end{align*}
\]

with \( \sum_{j,\delta_j = 1}^{} w^*_{iy} = 1 \), and \( \hat{\theta} \) being the MLE obtained from (2.4). The second line follows from the MAR assumption. Furthermore, we can write

\[
\begin{align*}
f(y_j|\delta_j = 1) &= \int f(y_j|x, \delta_j = 1)f(x|\delta_j = 1)dx \\
&= \int f(y_j|x)f(x|\delta_j = 1)dx \\
&\approx \frac{1}{N^r_{k=1}} \sum_k^r f(y_j|x_k),
\end{align*}
\]

where the second equality follows from the MAR assumption, and the last (approximate) equality follows by approximating the integral by the population empirical distribution, and \( N^r \) is the number of respondents in the population. Using the survey weights, we can approximate

\[
f(y_j|\delta_j = 1) \approx \frac{\sum_{k \in A_r} w_k f(y_j|x_k)}{\sum_{k \in A_r} w_k}
\]

and the fractional weights in (3.1) are computed from

\[
w^*_{iy} \propto \frac{f(y_j|x; \hat{\theta})}{\sum_{k \in A_r} w_k f(y_j|x_k; \hat{\theta})}
\]

with \( \sum_{k \in A_r} w^*_{iy} = 1 \). In (3.3), the point mass \( w^*_{iy} \) assigned to donor \( y_j \) for missing unit \( i \) is expressed by the ratio of the density \( f(y|x) \). Thus, for each missing unit \( i \), \( n^r = |A_r| \) observations are used as donors for the hot deck imputation using \( w^*_{iy} \) as the fractional weights. Such fractional imputation can be called full fractional imputation (FFI) because there is no randomness due to the imputation mechanism. The FFI estimator of \( \eta \), defined by \( \sum_{i,j}^N U(\eta; x_i, y_j) = 0 \), is then computed by solving

\[
\sum_{i \in A} w_i \left( \delta_i U(\eta; x_i, y_i) + (1-\delta_i) \sum_{j \in A_r} w^*_{ij} U(\eta; x_j, y_j) \right) = 0,
\]

\[
(3.4)
\]
where \( w^*_y \) is defined in (3.3). Note that the imputed estimating equation (3.4) is a good approximation to the expected estimating equation in (2.2).

In survey sampling, an imputed data set with a large imputation size may not be desirable. Thus, instead of taking all the observations in \( A_k \) as donors for each missing item, a subset of \( A_k \) can be selected to reduce the size of the donor set of missing \( y_i \). Thus, the selection of the donors is viewed as a sampling problem and we use an efficient sampling design and weighting techniques to obtain efficient imputation estimators. For the donor selection mechanism, efficient sampling designs, such as a stratified sampling design or systematic Proportional-to-Size (PPS) sampling, can be used to select donors of size \( m \). A systematic PPS sampling for fractional hot deck imputation can be described as follows:

1. Within each \( i \) with \( \delta_i = 0 \), sort the donors in the full respondent set \( \{ y_j; \delta_j = 1 \} \) in ascending order as \( y_{(1)} \leq \cdots \leq y_{(r)} \) and use \( w^*_y(j) \) to denote the fractional weight associated with \( y_{(j)} \). That is, \( w^*_y(j) = w_k^* \) for \( y_{(j)} = y_k \).

2. Partition \([0,1]\) by \( \{ I_j = \left[ \sum_{k=0}^{j-1} w^*_y(j), \sum_{k=0}^{j} w^*_y(j) \right], j = 1, \ldots, r-1 \} \), where \( w^*_{y(0)} = 0 \).

3. Generate \( u \sim \text{uniform}(0,1/m) \) and let \( u_k = u + k/m, \ k = 0, \ldots, m-1 \). For \( k = 0, \ldots, m-1 \), if \( u_k \in I_j \) for some \( 0 \leq j \leq r-1 \), include \( j \) in the sample \( D_i \).

After we select \( D_i \) from the complete set of respondents, the selected donors in \( D_i \) are assigned with the initial fractional weights \( w^*_{y(0)} = 1/m \). The fractional weights are further adjusted to satisfy

\[
\sum_{i \in \tilde{A}} w_i \left( 1 - \delta_i \right) \sum_{j \in \tilde{A}} w^*_y(j) q(x_i, y_j) = \sum_{i \in \tilde{A}} w_i \left( 1 - \delta_i \right) \sum_{j \in A_k} w^*_y(j) q(x_i, y_j), \tag{3.5}
\]

for some \( q(x, y_j) \), and \( \sum_{j \in \tilde{A}} w^*_y(j) = 1 \) for all \( i \) with \( \delta_i = 0 \), where \( w^*_y \) is the fractional weights for FFI method, as defined in (3.3). Regarding the choice of the control function \( q(x, y) \) in (3.5), we can use \( q(x, y) = \left( y, y^2 \right) \), which keeps the empirical distributions of \( y \) for \( D_i \) and \( A_k \) as close as possible in the sense that the first and second moment of \( y \) are the same. Other choices can also be considered. See Fuller and Kim (2005).

The problem of adjusting the initial weights to satisfy certain constraints is often called calibration and the resulting fractional weights can be called calibrated fractional weights. Using the idea of regression weighting, the final calibration fractional weights that satisfy (3.5) and \( \sum_{j \in \tilde{A}} w^*_y(j) = 1 \) can be computed by

\[
w^*_{y,k} = w_{y(0)}^* + w_{y(0)}^* \Delta \left( \tilde{q}_y^* - \tilde{q}_y^\dagger \right), \tag{3.6}
\]

where \( \tilde{q}_y^* = q(x, y_j) \), \( \tilde{q}_y^\dagger = \sum_{j \in \tilde{A}} w^*_{y(j)} q^*_y \),

\[
\Delta = \left\{ C_q - \sum_{i \in \tilde{A}} w_i \left( 1 - \delta_i \right) \sum_{j \in \tilde{A}} w^*_y(j) q^*_y \right\}^{-1} \left\{ \sum_{i \in \tilde{A}} w_i \left( 1 - \delta_i \right) \sum_{j \in A_k} w^*_y(j) \left( \tilde{q}_y^* - \tilde{q}_y^\dagger \right) \right\}.\]
and \[ C_q = \sum_{i \in A} w_i \left\{ (1 - \delta_i) \sum_{j \in d_i} w_j^* q(x_i, y_j) \right\} \]. Here, \( B^{02} \) denotes \( BB^T \). Some of the fractional weights computed by (3.6) can take negative values. If that happens, algorithms alternative to regression weighting should be used. For example, consider entropy weighting, where the fractional weights of the form

\[ w_{j, w}^* = \frac{w_j^* \exp(\Delta q_j^*)}{\sum_{k \in d_i} w_k^* \exp(\Delta q_k^*)} \]  

are approximately equal to the regression fractional weights in (3.6) and are always positive. Once the calibration fractional weights are obtained, the FHDI estimator of \( \eta \) is then computed by solving

\[ \sum_{i \in A} w_i \left\{ \delta U(\eta; x_i, y_i) + (1 - \delta_i) \sum_{j \in d_i} w_{j, w}^* U(\eta; x_i, y_j) \right\} = 0. \] (3.8)

For variance estimation, a replication method can be used. See Appendix A.1 for a brief discussion of the replication variance estimator for the proposed method.

Furthermore, the proposed method can handle non-ignorable non-response under the correct specification of the response model. See Appendix A.3 for the extension to non-ignorable non-response case.

### 4 Robustness

We now discuss the robustness of the proposed method against a small departure from the assumed parametric model. The robustness feature in our proposed estimator is defined to be robust against imputation model misspecification, a small exponential tilting of the true model. For simplicity of the presentation, assume that the sampling design is simple random sampling and the realized sample is a random sample from the superpopulation model.

We assume that the true model \( g(y \mid x) \) does not belong to \( \{ f(y \mid x; \theta) ; \theta \in \Omega \} \). However, we can still specify a working model \( f(y \mid x; \theta) \) and compute the MLE of \( \theta \). It is well known (White 1982) that the MLE converges to \( \theta^* \), the minimizer of the Kullback-Leibler information

\[ K(\theta) = E_g \left[ \log \left\{ \frac{g(Y \mid x)}{f(Y \mid x; \theta)} \right\} \right] \]

for \( \theta \in \Omega \). Sung and Geyer (2007) discussed the asymptotic properties of the Monte Carlo MLE of \( \theta \) under missing data.

To formally discuss robustness, suppose that the true distribution \( g(y \mid x) \) belongs to the neighborhood

\[ \mathcal{N}_\varepsilon = \left\{ g; D(g, f) < \frac{1}{2}\varepsilon^2 \right\} \] (4.1)

for some radius \( \varepsilon > 0 \), where

\[ D(g, f) = \int \log \left( \frac{g}{f} \right) g \, dy, \] (4.2)
is the Kullback-Leibler distance measure. The neighborhood (4.1) can be characterized in the following way. Let \( z(x,y,\theta) \) be a function of \( x, y \) and \( \theta \), standardized to satisfy \( E_{Y|x}(z) = 0 \) and \( Var_{Y|x}(z) = 1 \), and define

\[
g(y|x) = f(y|x;\theta) \exp\{\varepsilon z(x,y,\theta) - \kappa(x,\theta)\},
\]

where

\[
\kappa = \log \left( E_{Y|x}\left[ \exp\{\varepsilon z(x,Y,\theta)\} \right] \right).
\]

For small \( \varepsilon > 0 \) it can be shown that

\[
\kappa \equiv D(g,f) \equiv \frac{1}{2} \varepsilon^2.
\]

Equation (4.3) represents an extensive set of distributions close to \( f(y|x;\theta) \) created by varying \( z(x,y,\theta) \) over different standardized functions, where \( z \) and \( \varepsilon \) contain some geometric interpretation which represent the direction and magnitude of the misspecification respectively. For \( p \)-dimension parameter \( \theta \), we can specify the directions of the misspecification as

\[
\left( z_1, z_2, \ldots, z_p \right)^T = I_{\theta}^{-1/2} s(x,y,\theta),
\]

where \( s(x,y,\theta) = \partial \log f(y|x;\theta)/\partial \theta \) and \( I_{\theta} \) is the information matrix for \( \theta \). Represent \( z(x,y,\theta) \) as

\[
z(x,y,\theta) = \lambda^T I_{\theta}^{-1/2} s(x,y,\theta),
\]

where \( \sum_{i=1}^{p} \hat{\lambda}_i^2 = 1 \), then \( z(x,y,\theta) \) satisfies the standardization criterion of \( E_{Y|x}(z) = 0 \) and \( Var_{Y|x}(z) = 1 \).

See Copas and Eguchi (2001) for further discussion of this expression.

Let \( w_{y,g}^* \) be the fractional weight of the form (3.3) using the true density \( g \) and \( w_{y,f}^* \) be the corresponding fractional weight using the "working density" \( f \). By the special construction of the weights, we can establish

\[
w_{y,g}^* \approx w_{y,f}^* + \varepsilon \lambda^T I_{\theta}^{-1/2} \frac{\partial}{\partial \theta} \left( w_{y,f}^* \right).
\]

Proof of (4.5) is given in Appendix A.2. Thus

\[
\sum_i w_i \sum_j w_{y,g}^i U(\eta;x_i,y_j) \equiv \sum_i w_i \sum_j w_{y,f}^i U(\eta;x_i,y_j)
+ \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_i w_i \sum_j \frac{\partial}{\partial \theta} \left( w_{y,f}^i \right) U(\eta;x_i,y_j).
\]

For small \( \varepsilon \), we have

\[
\sum_i w_i \sum_j w_{y,g}^i U(\eta;x_i,y_j) \approx \sum_i w_i \sum_j w_{y,f}^i U(\eta;x_i,y_j),
\]

and so the resulting estimator of \( \eta \) from \( \sum_i w_i \sum_j w_{y,f}^i U(\eta;x_i,y_j) = 0 \) will be close to the true value \( \eta_0 \).
5 Simulation study

We performed two simulation studies. In Section 5.1, we compared the performance of the proposed method with some other imputation methods in a correctly specified model and a misspecified model, respectively, with ignorable missing data. In Section 5.2, we compared the statistical power of a test based on FHDI versus MI.

5.1 Simulation one

The first simulation study tested the performance of the proposed method under the setup of ignorable missing data. We used two sets of models to generate the observations. In model A, $y_i = 0.5x_i + e_i$, where $x_i \sim \exp(1)$, $e_i \sim N(0,1)$, with $x_i$ and $e_i$ being independent. In model B, $y_i = 0.5x_i + e_i$, where $x_i \sim \exp(1)$, $e_i \sim \chi^2(2) - 2 / 2$, with $x_i$ and $e_i$ being independent. Random samples of size $n = 200$ were separately generated from the two models. In addition to $(x_i, y_i)$, we also generated $\delta_i$ from Bernoulli($\pi_i$), where $\pi_i = \{1 + \exp(-0.2 - x_i)\}^{-1}$. Variable $x_i$ was always observed but variable $y_i$ was observed if and only if $\delta_i = 1$. The overall response rates were about 65% in both cases. We used $B = 2,000$ Monte Carlo samples in the simulation.

From each of the Monte Carlo samples, one generated from model A and the other generated from model B, we computed the following eight estimators:

1. Full sample estimator (Full) that is computed using the full sample.

2. Predictive Mean Matching (PMM) is a semi-parametric imputation method, which fills in a value randomly from observations that are closest to the predicted value obtained from $f(y | x)$. The PMM was implemented using "mice.impute.pmm" function in R.

3. Multiple imputation (MI) estimator with imputation size $m = 10$, where the imputed values are generated from the normal-theory regression model, as considered in Schenker and Welsh (1988).

4. Parametric fractional imputation (PFI) estimator without calibration with imputation size $m = 10$.

5. Parametric fractional imputation (PFI_cal) estimator with calibration with imputation size $m = 10$. The fractional weights are computed using the calibration method in (3.6) with $q = \{y, y^2\}$.

6. Full fractional imputation (FFI) estimator using the full set of respondents as imputation values, i.e. the imputation size $m = n_k$, where $n_k$ is the size of $A_k$.

7. Fractional hot deck imputation (FHDI) estimator without calibration using a small subset of respondents of size $m = 10$ as imputation values.
8. Fractional hot deck imputation (FHDI_cal) estimator with calibration using a small subset of respondents of size \( m = 10 \) as imputation values. The fractional weights are computed using the calibration method in (3.6) with \( \mathbf{q} = (y, y^2) \).

Multiple imputation is an approach of generating imputed values with simplified variance estimation. In this procedure, Bayesian methods of generating imputed values are considered, where \( m > 1 \) imputed values are generated from the posterior predictive distribution. Using the imputed values \( y^{*(1)}, \ldots, y^{*(m)} \), the multiple imputation estimator of \( \eta \), denoted by \( \hat{\eta}_{MI} \) is

\[
\hat{\eta}_{MI} = \frac{1}{m} \sum_{k=1}^{m} \hat{y}^{(k)}
\]

where \( \hat{y}^{(k)} \) is the complete response estimator applied to the \( k \)-th imputed data set. Rubin’s formula can be used for variance estimation in MI,

\[
\hat{V}_{MI} (\hat{\eta}_{MI}) = W_m + \left( 1 + \frac{1}{m} \right) B_m,
\]

where \( W_m = m^{-1} \sum_{k=1}^{m} \hat{V}^{(k)} \), \( B_m = (m - 1)^{-1} \sum_{k=1}^{m} \left( \hat{\eta}^{(k)} - \hat{\eta}_{MI} \right)^2 \), and \( \hat{V}^{(k)} \) is the variance estimator of \( \hat{\eta}^{(k)} \) under complete response applied to the \( k \)-th imputed data set.

In both models, we used the normal density with mean \( \beta_0 + \beta_1 x \) and variance \( \sigma^2 \) as the working model for imputation. Thus, the working model is the true model in model A but not true in model B.

We considered three parameters: \( \theta_1 = E(Y) \) the population mean of \( Y \), \( \theta_2 = Pr(Y < 1) \) the proportion of \( Y \) less than one, and \( \theta_3 \), the 0.5 quantile of \( Y \). In estimating \( \theta_2 \) under full sample, we used \( \hat{\theta}_{2,n} = n^{-1} \sum_{i=1}^{n} I(y_i < 1) \). In estimating \( \theta_3 \) under full sample, we used \( \hat{\theta}_{3,n} = \hat{F}^{-1}(p) = \inf \{ y : \hat{F}(y) > p \} \), where \( \hat{F}(y) = n^{-1} \sum_{i=1}^{n} I(y_i < y) \) and \( p = 0.5 \).

Table 5.1 and Table 5.2 show Monte Carlo means, standardized variance (Std Var) and standardized mean squared errors (Std MSE) of the eight estimators under model A and under model B, respectively. The standardized variance (mean squared error) is calculated as the ratio of variance (mean squared error) and the variance (mean squared error) of the full sample estimator multiplied by 100, which measures the increased variance (mean squared error) due to imputation relative to the full sample estimator. As for the Monte Carlo means (4th column), the imputation estimators are all unbiased for estimating \( \theta_1 \), \( \theta_2 \), and \( \theta_3 \) under model A. Under model B, PMM, MI, PFI, PFI_cal for estimating \( \theta_3 \) have much larger biases in absolute values than FFI, FHDI, and FHDI_cal under model misspecification in this simulation. Regarding the standardized variance and standardized mean squared error (5th and 6th column), PFI_cal is more efficient than FHDI. The reason is that in PFI, the imputed values are generated according to the conditional distribution \( f(y|x) \) directly; whereas in FHDI, the imputed values can be taken from respondents with dominantly large fractional weights. The effective imputation data size is determined by the imputed observations with large fractional weights, which also contribute to the loss of efficiency. FHDI loses efficiency in order to gain robustness. Lastly, FHDI with \( m = 10 \) has slightly larger standardized variance for \( \theta_2 \) than FFI, because of the additional variability due to the sampling procedure.

Comparing PFI with PFI_cal and FHDI with FHDI_cal, the calibration step improves the efficiency a little bit. The PMM shows the largest variance in all scenarios.
Table 5.1
Monte Carlo mean, standardized variance and standardized mean squared error of point estimators in Model A of Simulation one.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>Std Var</th>
<th>Std MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_y$</td>
<td>Full</td>
<td>0.50</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.50</td>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.50</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.50</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.50</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.50</td>
<td>156</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.50</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.50</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>$Pr(Y &lt; 1)$</td>
<td>Full</td>
<td>0.68</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.68</td>
<td>168</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.68</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.68</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.68</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.68</td>
<td>137</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.68</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.68</td>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>Quantile</td>
<td>Full</td>
<td>0.47</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.47</td>
<td>184</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.47</td>
<td>111</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.47</td>
<td>111</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.47</td>
<td>111</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.47</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.47</td>
<td>142</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.47</td>
<td>141</td>
<td>141</td>
</tr>
</tbody>
</table>

Table 5.2
Monte Carlo mean, standardized variance and standardized mean squared error of point estimators in Model B of Simulation one.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>Std Var</th>
<th>Std MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_y$</td>
<td>Full</td>
<td>0.50</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.50</td>
<td>172</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.50</td>
<td>131</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.50</td>
<td>131</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.50</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.50</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.50</td>
<td>147</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.50</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>$Pr(Y &lt; 1)$</td>
<td>Full</td>
<td>0.75</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.75</td>
<td>166</td>
<td>166</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.75</td>
<td>140</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.75</td>
<td>138</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.75</td>
<td>137</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.75</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.75</td>
<td>145</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.75</td>
<td>140</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>Quantile</td>
<td>Full</td>
<td>0.26</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMM</td>
<td>0.24</td>
<td>191</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI ($m = 10$)</td>
<td>0.31</td>
<td>122</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PMI ($m = 10$)</td>
<td>0.31</td>
<td>123</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI ($m = 10$)</td>
<td>0.31</td>
<td>122</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FPI cal ($m = 10$)</td>
<td>0.31</td>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD ($m = 10$)</td>
<td>0.26</td>
<td>144</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHD cal ($m = 10$)</td>
<td>0.26</td>
<td>139</td>
<td>139</td>
</tr>
</tbody>
</table>
For variance estimation, we considered replication variance estimation for FFI and FHDI, particularly the delete-1 Jackknife variance estimation, which is described in Appendix A.1. We also considered variance estimation in MI, which uses Rubin's formula (5.1).

Table 5.3 shows the Monte Carlo relative biases of the variance estimators, which is calculated as $\left[ E_{MC}\{\hat{V}\} - V_{MC}\{\hat{\theta}\} \right] / V_{MC}\{\hat{\theta}\}$, where $E_{MC}\{\hat{V}\}$ is the Monte Carlo mean of variance estimates $\hat{V}$, and $V_{MC}\{\hat{\theta}\}$ is the Monte Carlo variance of the point estimates $\hat{\theta}$. The relative bias of the variance estimator in FFI and FHDI is reasonably small for all parameters considered in both models, suggesting that the replication variance estimator is valid. The relative bias and $t$-statistics of variance estimator in MI are small for $\theta_1$ but quite large for $\theta_2$ even when the working model is true (model A). Rubin's formula is based on the following decomposition,

$$V(\hat{\theta}_M) = V(\hat{\theta}_n) + V(\hat{\theta}_m - \hat{\theta}_n),$$

(5.2)

where $\hat{\theta}_n$ is the full sample estimator of $\eta$. Basically, the $W_m$ term in (5.1) estimates $V(\hat{\theta}_n)$ and the $(1 + m^{-1})B_m$ term in (5.1) estimates $V(\hat{\theta}_m - \hat{\theta}_n)$. The decomposition (5.2) holds when $\hat{\theta}_n$ is the MLE of $\theta$, which is the congeniality condition of $\hat{\theta}_n$ (Meng 1994). For general case, we have

$$V(\hat{\theta}_M) = V(\hat{\theta}_n) + V(\hat{\theta}_m - \hat{\theta}_n) + 2\text{Cov}(\hat{\theta}_m - \hat{\theta}_n, \hat{\theta}_n)$$

(5.3)

and Rubin's variance estimator can be biased if $\text{Cov}(\hat{\theta}_m - \hat{\theta}_n, \hat{\theta}_n) \neq 0$. The congeniality condition holds true for estimating the population mean; however, it does not hold for the method of moments estimator of $Pr(Y < 1)$. Note that the imputed estimator of $\theta_2 = Pr(Y < 1)$ can be expressed as

$$\hat{\theta}_{2,i} = n^{-1} \sum_{i=1}^{n} \left[ \delta I(y_i < 1) + (1-\delta)E\{I(y < 1)|x_i; \hat{\mu}, \hat{\sigma}\} \right].$$

(5.4)

Thus, the imputed estimators of $\theta_2$ "borrows strength" by making use of extra information associated with $f(y|x)$. That is, the normality of $f(y|x)$ is used in computing the conditional expectation in (5.4), which improves the efficiency of the imputed estimator for $\theta_2$. The same phenomenon also holds for $\theta_3$. In Table 5.1, the increase of variance due to imputation for MI with $m = 10$ is about 35% for $\theta_1$ but only 12% and 11% for $\theta_2$ and $\theta_3$, respectively, which shows the phenomenon of "borrowing strength" for estimating $\theta_2$ and $\theta_3$ thanks to the use of extra information in the imputation stage. Thus, when the congeniality conditions do not hold, the imputed estimator improves the efficiency but Rubin's variance estimator does not recognize this improvement.
Table 5.3
Monte Carlo relative bias of the replication variance estimator in Simulation one.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Method</th>
<th>R.B. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>-2.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>-0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>-0.80</td>
</tr>
<tr>
<td>*A</td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>8.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>-5.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>-5.12</td>
</tr>
<tr>
<td></td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>19.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>4.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>3.78</td>
</tr>
<tr>
<td>*B</td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>-0.56</td>
</tr>
<tr>
<td></td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>-3.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>-1.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>-3.25</td>
</tr>
<tr>
<td></td>
<td>( \hat{V}(\theta) )</td>
<td>MI (( m = 10 ))</td>
<td>-8.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FFI (( m = n_x ))</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FHDI_cal (( m = 10 ))</td>
<td>3.80</td>
</tr>
</tbody>
</table>

5.2 Simulation two

Simulation two tested the power of the proposed method in a hypothesis test using the null model as the imputation model. Samples of bivariate data \((x_i, y_i)\) of size \( n = 100 \) were generated from

\[
y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i^2 - 1) + \epsilon_i
\]

(5.5)

where \((\beta_0, \beta_1, \beta_2) = (0, 0.9, 0.06), x_i \sim N(0,1), \epsilon_i \sim N(0,0.16)\), with \(x_i\) and \(\epsilon_i\) being independent. The variable \(x_i\) is always observed but the probability that \(y_i\) responds is 0.5. Monte Carlo samples were generated independently for \(B = 10,000\) times. We are interested in testing \(H_0: \beta_2 = 0\) from the respondents. We compared FHDI with MI using the same imputation size \(m = 30\). The imputation model is the null model,

\[
y_i = \beta_0 + \beta_1 x_i + \epsilon_i.
\]

That is, the imputation model uses extra information of \(\beta_2 = 0\). From the imputed data, we fit model (5.5) and computed the power of a test \(H_0: \beta_2 = 0\) at the significant level of 0.05. In addition, we also considered the complete case (CC) method that only uses the respondents for regression.

Table 5.4 shows the Monte Carlo mean and variance of the point estimators, relative bias of the variance estimator and the Monte Carlo power of testing \(H_0: \beta_2 = 0\). In each Monte Carlo sample, we constructed a 95% Wald confidence interval of \(\beta_2\) as \((\hat{\beta}_2 - 1.96\hat{V}^{1/2}, \hat{\beta}_2 + 1.96\hat{V}^{1/2})\) and reject the null hypothesis if \(\beta_2 = 0\) does not fall in the Wald confidence interval. The Monte Carlo power is calculated as...
the relative frequency of rejecting the null hypothesis among the Monte Carlo samples. From the second column, FHDI and MI estimators are biased for $\beta_2$, as expected since in imputation the imputation model is the null model and it is slightly different from the true model that generated sample. The bias of FHDI is smaller than that of MI because of the robustness of FHDI discussed in Section 4. In MI, 50% of the imputed MI data comes from the null model and the other 50% from the true model, so the slope $\beta_2$ is attenuated to zero by half of the true slope. In FHDI, though we used the null model to calculate the fractional weights, the imputed data come from the true model which reduces the bias. Moreover, MI provides more efficient point estimators than the CC method but variance estimation is very conservative (about 180% overestimation). Because of the serious positive bias of MI variance estimator, the statistical power of the test based on MI is actually lower than the CC method. On the other hand, FHDI also provides more efficient point estimators than the CC method and variance estimation is essentially unbiased, the statistical power of the test based on FHDI is higher than the CC method.

<table>
<thead>
<tr>
<th>Table 5.4</th>
<th>Simulation results based on 10,000 Monte Carlo samples in Simulation two.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>$E(\hat{\beta}_2)$</td>
</tr>
<tr>
<td>FHDI</td>
<td>0.046</td>
</tr>
<tr>
<td>MI</td>
<td>0.028</td>
</tr>
<tr>
<td>CC</td>
<td>0.060</td>
</tr>
</tbody>
</table>

6 Concluding remarks

We have proposed a fractional hot deck imputation method that uses a parametric model for $f(y | x)$ when $x$ contains continuous components. The proposed method provides robust estimation for the parameters in the sense that the imputation model is not necessarily equal to the data-generating model. The price we pay in the FHDI is the loss of efficiency in point estimation. Under our first simulation, the FHDI estimator for $P(Y < 1)$ has the second largest variance but the smallest mean squared error when the working model is not true, as compared with other estimators.

The loss of efficiency mainly comes from the fact that the fractional weights are more variable than those under the PFI method because some of $x_j$ are not useful in imputing $y_j$. That is, the value of $f(y_j | x_j; \hat{\theta})$ can be very small. The fractional hot deck imputation under a small imputation size (e.g. $m = 10$) does not increase the variance significantly, as can be seen in Table 5.1 under model A.

The proposed fractional imputation method can actually be used to develop a single imputation method by applying FHDI with $m = 1$, which selects an imputed value with probability proportional to the fractional weight for each missing unit. In this case, the FHDI can be used to develop a single imputation that is still robust against model misspecification. However, weighting calibration cannot co-exist with single imputation. Calibration constraints can still be achieved by employing the balanced imputation method as discussed in Chauvet, Deville and Haziza (2011) or the rejective Poisson sampling of Fuller (2009). Further investigation along this direction will be a topic of future research.
Acknowledgements

We thank two anonymous referees and the associate editor for very helpful comments. This research was partially supported by a grant from NSF (MMS-121339) and by the Cooperative Agreement between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University.

Appendix

A.1 Replication variance estimation

For variance estimation, replication methods can be used. Let \( w_{ik} \) be the \( k \)-th replication weights such that

\[
\hat{V}_{rep} = \sum_{k=1}^{L} c_k (\hat{Y}^{[k]} - \hat{Y})^2
\]

is consistent for the variance of \( \hat{Y} = \sum_{i \in A} w_i y_i \), where \( L \) is the replication size, \( c_k \) is the \( k \)-th replication factor that depends on the replication method and the sampling mechanism, and \( \hat{Y}^{[k]} = \sum_{i \in A} w_{ik} y_i \) is the \( k \)-th replicate of \( \hat{Y} \). In delete-1 jackknife variance estimation, \( L = n \) and \( c_k = (n-1)/n \).

To apply the replication method in FFI, we first apply the replication weights \( w_{ik} \) in (2.4) to compute \( \hat{\theta}^{[k]} \). Once \( \hat{\theta}^{[k]} \) is obtained, we use the same imputed values to compute the initial replication fractional weights

\[
w^{*}_{ij} \propto w_{ij}^{[k]} f(y_j | x_j; \hat{\theta}^{[k]}) \left/ \left( \sum_{j \in A} f(y_j | x_j; \hat{\theta}^{[k]}) \right) \right.,
\]

(A.1)

with \( \sum_{j \in A} w^{*}_{ij} = 1 \). The variance of \( \hat{\eta}_{FFI} \), computed from (3.4), is then computed by

\[
\hat{V}_{rep} = \sum_{k=1}^{L} c_k (\hat{\eta}_{FFI}^{[k]} - \hat{\eta}_{FFI} )^2,
\]

where \( \hat{\eta}_{FFI}^{[k]} \) comes from solving

\[
\sum_{i \in A} w_{ij}^{[k]} \left\{ \delta U(\eta; x_i, y_i) + (1 - \delta_i) \sum_{j \in A} w_{ij}^{*} U(\eta; x_j, y_j) \right\} = 0,
\]

and \( w^{*}_{ij} \) is defined in (A.1).

We now discuss replication variance estimation of the FHDI estimator \( \hat{\eta}_{FHDI} \) computed from (3.8). Define \( d_{ij} = 1 \) if \( j \in D_i \) and \( d_{ij} = 0 \) otherwise. Note that \( \hat{\eta}_{FHDI} \) is computed via two steps: in the first step, a systematic PPS sampling is used with the selection probability proportional to the fractional weights from the FFI method. In the second step, the calibration weighting method using the constraint (3.5) with
\[ \sum_{j \in D_j} d_{ij} w_{ij} = 1 \] is used. Thus, the replicate fractional weights are also computed in two steps. Firstly, the initial replication fractional weight for \( w_{ij0} = 1/m \) is then given by
\[
\frac{d_{ij} \left( w_{ij}^{k(k)} / w_{ij} \right)}{\sum_{i \in A} d_{ij} \left( w_{ij}^{k(k)} / w_{ij} \right)}.
\] (A.2)

where \( w_{ij} \) is the fractional weight for FFI defined in (2.6) and \( w_{ij}^{k(k)} \) is the \( k \)-th replication fractional weight for FFI defined in (A.1). Secondly, the replication fractional weights are adjusted to satisfy the calibration constraints. The calibration equation for replication fractional weights corresponding to (3.5) is then
\[
\sum_{i \in A} w_{ij}^{k(k)} \left( 1 - \delta_i \right) \sum_{j \in D_j} w_{ij}^{k(k)} q(x_i, y_j) = \sum_{i \in A} w_{ij}^{k(k)} \left( 1 - \delta_i \right) \sum_{j \in D_j} w_{ij}^{k(k)} q(x_i, y_j).
\] (A.3)

and \( \sum_{j \in D_j} w_{ij}^{k(k)} = 1 \). Either regression weighting or entropy weighting can be used to obtain the replication fractional weights satisfying the constraints. Once the replicate fractional weights are obtained, the replicate estimate \( \hat{\eta}^{[k]} \) is computed by solving
\[
\sum_{i \in A} w_{ij}^{k(k)} \left( \delta_i U(\eta; x_i, y_j) + (1 - \delta_i) \sum_{j \in D_j} w_{ij}^{k(k)} U(\eta; x_i, y_j) \right) = 0.
\]

The replication variance estimator of \( \hat{\eta} \), computed from (3.8), is given by
\[
\hat{V}_{rep}(\hat{\eta}) = \sum_{k=1}^{K} c_k (\hat{\eta}^{[k]} - \hat{\eta})^2.
\]

Because \( \hat{\eta} \) is a smooth function of \( \hat{\theta} \), the consistency of \( \hat{V}_{rep}(\hat{\eta}) \) follows directly from the standard argument of the replication variance estimation (Shao and Tu 1995).

**A.2 Proof of Equation (4.5)**

Using
\[
g(y_j | x_j) = \frac{f(y_j | x_j)}{f(y_j | x_k)} \exp(\Delta_{kj} - \kappa(x_j) + \kappa(x_k))
\]
where \( \Delta_{kj} = z(x_j, y_j; \theta) - z(x_k, y_j; \theta) \). Based on Taylor linearization and the fact of (4.4), we have
\[
\frac{g(y_j | x_j)}{g(y_j | x_k)} \approx \frac{f(y_j | x_j)}{f(y_j | x_k)} \left( 1 + \Delta_{kj} \right).
\]

If we know the true density, the correct fractional weights in (3.3) can be expressed by
\[ w^*_{y,x} \propto \frac{g(y_j | x_j)}{\sum_{k \delta_j = 1} w_k g(y_j | x_k)} \]

\[ \propto \frac{1}{\sum_{k \delta_j = 1} w_k \left[ \frac{g(y_j | x_k)}{g(y_j | x_j)} \right]} \]

\[ \propto \frac{1}{\sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \exp \left( \varepsilon \Delta_{k,j} - \kappa(x_j) + \kappa(x_k) \right)} \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \left( 1 + \varepsilon \Delta_{k,j} \right) \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] + \varepsilon \sum_{k \delta_j = 1} w_k \left[ \frac{z(x_k, y_j) - z(x_j, y_j)}{f(y_j | x_j)} \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] + \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \left( s(x_k, y_j) - s(x_j, y_j) \right) \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] + \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{\partial}{\partial \theta} \left[ f(y_j | x_k) \right] \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \left[ 1 - \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \left[ 1 - \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \left[ 1 - \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \right] \]

\[ \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \left[ 1 - \varepsilon \lambda^T I_{\theta}^{-1/2} \sum_{k \delta_j = 1} w_k \left[ \frac{f(y_j | x_k)}{f(y_j | x_j)} \right] \right] \]

\[ = a_\theta + \varepsilon \lambda^T I_{\theta}^{-1/2} \frac{\partial}{\partial \theta} a_\theta, \]
where \( a_y = f(y_j | x_i) / \sum_{k \in \delta_i \setminus i} w_k f(y_j | x_k) \) and \( a_i = \sum_{j \in \delta_i} a_{ij} \). So, \( w_{y,j}^* = a_y / a_i \) and
\[
 w_{y,j}^* \approx a_y + \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} a_y a_i + \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} a_i
\]
\[
 = a_y \left( 1 + \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} a_y \right) \left( a_i + \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} a_i \right)
\]
\[
 = a_y \left( 1 + \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} \log a_y \right) \left( 1 - \varepsilon \lambda^T I_0^{1/2} \frac{\partial}{\partial \theta} \log a_i \right)
\]
which proves (4.5).

**A.3 Extension to a non-ignorable missing case**

We consider an extension of the proposed method to a non-ignorable missing case. Under the non-ignorable missing assumption, both the conditional model \( f(y | x) \) and the response probability model \( P(\delta = 1 | x, y) \) are needed to evaluate the expected estimating function in (4.6). Let the response probability model be given by \( P(\delta = 1 | x, y) = \pi(x, y; \phi) \), for some \( \phi \) with a known \( \pi(\cdot) \) function. We assume that the parameters are identifiable as discussed in Wang, Shao and Kim (2013).

In PFI, according to Kim and Kim (2012), the MLE \( \hat{\theta}, \hat{\phi} \) can be obtained by solving
\[
 \sum_{i \in A} w_i \left\{ \delta_i S(\theta; x_i, y_i) + (1 - \delta_i) \sum_{j=1}^{m} w_{y,j}^* S(\theta; x_i, y_i^{(j)}) \right\} = 0, \tag{A.4}
\]
and
\[
 \sum_{i \in A} w_i \left\{ \delta_i S(\phi; x_i, y_i) + (1 - \delta_i) \sum_{j=1}^{m} w_{y,j}^* S(\phi; x_i, y_i^{(j)}) \right\} = 0, \tag{A.5}
\]
where \( S(\theta; x, y) = \partial \log f(y | x; \theta) / \partial \theta \), \( S(\phi; x, y) = \partial \log \pi(x, y; \phi) / \partial \phi \), and the fractional weights are given by
\[
 w_{y,j}^*(\theta, \phi) = \frac{f(y_i^{(j)} | x_i; \theta) \left[ 1 - \pi(x_i, y_i^{(j)}; \phi) \right] / h(y_i^{(j)} | x_i)}{\sum_{k=1}^{m} f(y_i^{(k)} | x_i; \theta) \left[ 1 - \pi(x_i, y_i^{(k)}; \phi) \right] / h(y_i^{(k)} | x_i)}. \tag{A.6}
\]
The solution to (A.4) and (A.5) can be obtained via the EM algorithm. In the EM algorithm, the E-step computes the fractional weights in (A.6) using the current parameter values and the M-step updates the parameter value \( \hat{\theta}^{(t)} \) and \( \hat{\phi}^{(t)} \) by solving

\[
\sum_{j \in A} w_j^* \left[ \delta_j S(\theta; x_i, y_i) + (1 - \delta_j) \sum_{j=1}^{m} w_j^* \left( \hat{\theta}^{(t)}, \hat{\phi}^{(t)} \right) S(\theta; x_i, y_i^{*^{(j)}}) \right] = 0,
\]

and

\[
\sum_{j \in A} w_j^* \left[ \delta_j S(\phi; x_i, y_i) + (1 - \delta_j) \sum_{j=1}^{m} w_j^* \left( \hat{\theta}^{(t)}, \hat{\phi}^{(t)} \right) S(\phi; x_i, y_i^{*^{(j)}}) \right] = 0.
\]

In the proposed FFI method, the fractional weights are given by

\[
w_j^* \propto f(y_j \mid x_i, \delta_j = 0; \theta, \phi) / f(y_j \mid \delta_j = 1)
\]

\[
\propto f(y_j \mid x_i, \theta) \left[ 1 - \pi(x_i, y_j; \phi) \right] / f(y_j \mid \delta_j = 1),
\]

with \( \sum_{j=1}^{m} w_j^* = 1 \). Because

\[
f(y_j \mid \delta_j = 1) = \int \pi(x, y_j) f(y_j \mid x) f(x) dx
\]

\[= \sum_{k \in A} w_k \pi(x_k, y_j) f(y_j \mid x_k). \tag{A.7}\]

The fractional weights can be computed from

\[
w_j^* \propto f(y_j \mid x_i; \theta) \left[ 1 - \pi(x_i, y_j; \phi) \right] \sum_{k \in A} w_k \pi(x_k, y_j; \phi) f(y_j \mid x_k; \theta), \tag{A.8}\]

with \( \sum_{j \in A} w_j^* = 1 \).

Thus, we can use the following EM algorithm to obtain the desired parameter estimates.

**I-step** For each missing unit \( i \in A_m = \{ i \in A; \delta_i = 0 \}, \) take \( m \) imputed values as \( y_i^{(1)}, \ldots, y_i^{(m)} \) from \( A_r \), where \( m = r \).

**E-step** The fractional weights are given by

\[
w_j^{*^{(i)}} \propto f(y_j \mid x_i, \hat{\theta}^{(i)}) \left[ 1 - \pi(x_i, y_j; \hat{\phi}^{(i)}) \right] \sum_{k \in A} w_k \pi(x_k, y_j; \hat{\phi}^{(i)}) f(y_j \mid x_k; \hat{\theta}^{(i)})
\]

and \( \sum_{j=1}^{m} w_j^{*^{(i)}} = 1 \).

**M-step** Update the parameter \( \hat{\theta}^{(t^{(i)}+1)} \) and \( \hat{\phi}^{(t^{(i)}+1)} \) by solving the following imputed score equations,
\[
\sum_{i \in A} w_i \left\{ \delta_i S(\Theta; x_i, y_i) + \left(1 - \delta_i \right) \sum_{j \in Ag} w_{ij}^{(n)} S(\theta; x_i, y_j) \right\} = 0,
\]
and

\[
\sum_{i \in A} w_i \left\{ \delta_i S(\phi; x_i, y_i) + \left(1 - \delta_i \right) \sum_{j \in Ag} w_{ij}^{(n)} S(\phi; x_i, y_j) \right\} = 0.
\]

Note that the I-step does not have to be repeated in the EM algorithm. Once the final parameter estimates are computed, the fractional weights are computed by (A.8), which serve as the selection probabilities for FHDI with a small imputation size \(m\). The same systematic PPS sampling method as discussed in Section 3 can be used to obtain FHDI.

References


