A simulation study for parametric fractional imputation
Jae Kwang Kim

1 Simulation Setup

• In the simulation study, \( B = 2,000 \) simulation samples of size \( n = 200 \) were generated with

\[
\begin{align*}
x_i & \sim N(2, 1) \\
y_{1i} | x_i & \sim N(\beta_0 + \beta_1 x_i, \sigma_{ee}) \\
y_{2i} | (x_i, y_{1i}) & \sim \text{Bernoulli}(p_i) \\
\delta_{1i} | (x_i, y_{1i}, y_{2i}) & \sim \text{Bernoulli}(\pi_{1i}) \\
\delta_{2i} | (x_i, y_{1i}, y_{2i}) & \sim \text{Bernoulli}(0.7)
\end{align*}
\]

where \((\beta_0, \beta_1) = (1, 0.7), \sigma_{ee} = 1, \)

\[
p_i = \frac{\exp(\phi_0 + \phi_1 x_i + \phi_2 y_{1i})}{1 + \exp(\phi_0 + \phi_1 x_i + \phi_2 y_{1i})}
\]

and

\[
\pi_{1i} = \frac{\exp(\psi_0 + \psi_1 x_i)}{1 + \exp(\psi_0 + \psi_1 x_i)}
\]

with \((\phi_0, \phi_1, \phi_2) = (-3, 0.5, 0.7)\) and \((\psi_0, \psi_1) = (0, 0.5)\). Variables \(x_i\) is always observed but \(y_{1i}\) and \(y_{2i}\) are subject to missingness. Variable \(\delta_{1i} = 1\) if \(y_{1i}\) is observed and \(\delta_{1i} = 0\) if \(y_{1i}\) is missing. Variable \(\delta_{2i} = 1\) if
$y_{2i}$ is observed and $\delta_{2i} = 0$ if $y_{2i}$ is missing. The average response rates for $y_{1i}$ and $y_{2i}$ are about 71% and 70%, respectively.

- We are interested in the following parameters:
  1. $\mu_1$: the marginal mean of $y_1$
  2. $\mu_2 = Pr (Y_2 = 1)$: the marginal mean of $y_2$
  3. $\beta$: the slope for the regression of $y_1$ on $x$.
  4. $\theta = Pr (Y_1 \leq 3)$: the proportion that $y_1$ is less than 3.

2 Fractional imputation

Procedures for PFI method:

[Step 0] Initial Imputation: Generate fractionally imputed values using $x$-information only.

  (a) Fit a parametric model $f_1(y_1 \mid x, \theta_1)$ for the conditional distribution of $y_1$ given $x$ among respondents of $y_1$. In this simulation setup, we use the following model

  $$y_{1i} \mid (x_i, \delta_{1i} = 1) \sim N(\beta_0 + \beta_1 x_i, \sigma_{ee})$$

  for some $\theta_1 = (\beta_0, \beta_1, \sigma_{ee})$.

  (b) Estimate parameter $\theta_1$ using the sample elements with $\delta_{1i} = 1$ only.

  (c) For each unit $i$ with $\delta_{1i} = 0$, generate $M = 100$ imputed values of $y_i$, say $y_{1i}^{*(1)}, \ldots, y_{1i}^{*(M)}$, from the estimated density $f_1(y_1 \mid x_i, \hat{\theta}_1)$. 

2
That is,
\[ y_{1i}^* \sim N \left( \hat{\beta}_0 + \hat{\beta}_1 x_i, \hat{\sigma}_{ee} \right). \]

(d) Similarly, fit a parametric model \( f_2(y_2 \mid x, y_1, \theta_2) \) for the conditional distribution of \( y_2 \) given \( x \) and \( y_1 \) among \( y_2 \)-respondents. In this simulation step, \( f_2(y_2 \mid x, y_1, \theta_2) \) follows a Bernoulli distribution with the probability
\[
p(x, y_1; \theta_2) = \frac{\exp(\phi_0 + \phi_1 x + \phi_2 y_1)}{1 + \exp(\phi_0 + \phi_1 x + \phi_2 y_1)}
\]
where \( \theta_2 = (\phi_0, \phi_1, \phi_2) \).

(e) Estimate parameter \( \theta_2 \) using the samples with \( \delta_{2i} = 1 \) only. That is, the parameter estimate for \( \theta_2 = (\phi_0, \phi_1, \phi_2) \) can be computed by solving
\[
\sum_{\delta_{1i}=1} \delta_{2i} \{ y_{2i} - p(x_i, y_{1i}; \theta_2) \} (1, x_i, y_{1i})'
\]
\[
+ M^{-1} \sum_{\delta_{1i}=0} \sum_{j=1}^M \delta_{2i} \{ y_{2i} - p(x_i, y_{1i}^*(j); \theta_2) \} (1, x_i, y_{1i}^*(j))' = 0.
\]
The solution can be easily obtained by modifying the existing software using weight \( 1/M \) for \( \delta_{1i} = 0 \).

(f) For each unit \( i \) with \( \delta_{2i} = 0 \), generate \( M = 100 \) imputed values of \( y_{2i} \) by
\[
y_{2i}^*(j) \sim \begin{cases} f_2(y_{2i} \mid x_i, y_{1i}, \hat{\theta}_2) & \text{if } \delta_{1i} = 1 \\ f_2(y_{2i} \mid x_i, y_{1i}^*(j), \hat{\theta}_2) & \text{if } \delta_{1i} = 0. \end{cases}
\]

[Step 1] **Fractional weighting (E-step):** For the current parameter estimate \( \hat{\theta}_{(t)} = \left( \hat{\beta}_{1(t)}, \hat{\beta}_{2(t)} \right)' \), where \( \hat{\theta}_{1(t)} = (\hat{\beta}_{0(t)}, \hat{\beta}_{1(t)}, \hat{\sigma}_{ee(t)})' \), \( \hat{\theta}_{2(t)} = (\hat{\phi}_{0(t)}, \hat{\phi}_{1(t)}, \hat{\phi}_{2(t)})' \), compute the fractional weights associated with the
imputed values. The fractional weights associate with \((x_i, y_{1i}^*, y_{2i}^*, z_i)\)
are
\[
w_{i(t)}^{j} \propto \frac{f_1 \left( y_{1i}^* \mid x_i, \hat{\theta}_{1(t)} \right) f_2 \left( y_{2i}^* \mid x_i, y_{1i}^*, \hat{\theta}_{2(t)} \right)}{f_1 \left( y_{1i}^* \mid x_i, \hat{\theta}_{1(0)} \right) f_2 \left( y_{2i}^* \mid x_i, y_{1i}^*, \hat{\theta}_{2(0)} \right)},
\]
where \(f_1 (y_1 \mid x, \theta_1)\) denotes the density for the conditional distribution
of \(y_1\) given \(x\) (which is normal in this case) and
\[
f_2 (y_2 \mid x, y_1, \theta_2) = \begin{cases} Pr (y_2 = 1 \mid x, y_1, \theta_2) & \text{if } y_2 = 1 \\ Pr (y_2 = 0 \mid x, y_1, \theta_2) & \text{if } y_2 = 0. \end{cases}
\]
In (2), it is understood that \(y_{1i}^* = y_{1i}\) if \(\delta_{1i} = 1\) and \(y_{2i}^* = y_{2i}\) if \(\delta_{2i} = 1\).

[Step 2] **Update parameter estimates (M-step):** Using the current fractional weights, compute the maximum likelihood estimator to update \(\hat{\theta}_{(t+1)}\) by solving the following imputed score equations:
\[
\bar{S}_{1(t)} (\theta_1) \equiv \sum_{i=1}^{n} \sum_{j=1}^{M} w_{i(t)}^{j} \left( y_{1i}^* - \beta_0 - \beta_1 x_i \right) (1, x_i) = (0, 0)
\]
\[
\bar{S}_{2(t)} (\theta_1) \equiv \sum_{i=1}^{n} \sum_{j=1}^{M} w_{i(t)}^{j} \left\{ \left( y_{1i}^* - \beta_0 - \beta_1 x_i \right)^2 - \sigma_{ee} \right\} = 0
\]
\[
\bar{S}_{3(t)} (\theta_2) \equiv \sum_{i=1}^{n} \sum_{j=1}^{M} w_{i(t)}^{j} \left\{ y_{2i}^* - p \left( x_i, y_{1i}^*; \psi_0, \psi_1, \psi_2 \right) \right\} \left( 1, x_i, y_{1i}^* \right) = (0, 0, 0)
\]

Newton-Raphson method is used to compute the maximum likelihood estimates.

[Step 3] Iteratively compute the parameters until convergence.

Using the fractionally imputed data, compute the point estimators and variance estimators (the details are described in the paper.)
3 Multiple imputation

[Step 0] Initial parameter generation & imputation:

(a) Fit a parametric model $f_1(y_1 \mid x, \theta_1)$ for the conditional distribution of $y_1$ given $x$ among respondents of $y_1$, where $\theta_1 = (\beta_0, \beta_1, \sigma_{ee})$.

(b) For each $k = 1, 2, \cdots, M$, generate $\theta_1^{(k)} = \left(\beta_0^{(k)}, \beta_1^{(k)}, \sigma_{ee}^{(k)}\right)$ by

$$\sigma_{ee}^{(k)} \sim r_1 \hat{\sigma}_{ee} / \chi^2(r_1)$$

$$\begin{pmatrix} \beta_0^{(k)} \\ \beta_1^{(k)} \end{pmatrix} \sim N \left( \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^{n} \delta_{1i} & \sum_{i=1}^{n} \delta_{1i} x_i \\ \sum_{i=1}^{n} \delta_{1i} x_i & \sum_{i=1}^{n} \delta_{1i} x_i^2 \end{pmatrix} \right)^{-1} \sigma_{ee}^{(k)},$$

where $r_1 = \sum_{i=1}^{n} \delta_{1i}$ and $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_{ee})$ satisfies

$$\sum_{i=1}^{n} \delta_{1i} \{y_i - \beta_0 - \beta_1 x_i\} (1, x_i)' = (0, 0)'$$

and

$$\sum_{i=1}^{n} \delta_{1i} \{(y_i - \beta_0 - \beta_1 x_i)^2 - \sigma_{ee}\} = 0.$$

(c) For $\delta_{1i} = 0$, generate the $k$-th imputed values of $y_{1i}$, denoted by $y_{1i}^{(k)}$, from $N(\beta_0^{(k)} + \beta_1^{(k)} x_i, \sigma_{ee}^{(k)})$. If $\delta_{1i} = 1$, then $y_{1i}^{(k)} = y_{1i}$.

(d) Also, for each $k = 1, 2, \cdots, M$, generate $\theta_2^{(k)} = (\phi_0^{(k)}, \phi_1^{(k)}, \phi_2^{(k)})'$ from a normal distribution with mean $\left(\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2\right)'$ and variance

$$\left(\sum_{i=1}^{n} \delta_{2i} S_i(\hat{\theta}_2) S_i(\hat{\theta}_2) \right)^{-1},$$

where $\hat{\theta}_2 = (\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2)$ satisfies

$$\sum_{i=1}^{n} \delta_{2i} S_i(\hat{\theta}_2) = (0, 0, 0)$$

with

$$S_i(\hat{\theta}_2) = \left\{ y_{2i} - p \left( x_i, y_{1i}^{(k)}, \hat{\theta}_2 \right) \right\} (1, x_i, y_{1i}^{(k)}).$$
Here, $p(x_i, y_{i1}; \theta_2)$ is defined in (1) and $y_{i1}^{* (k)}$ is the imputed value from (c).

(e) For $\delta_{2i} = 0$, generate the $k$-th imputed values of $y_{2i}$, denoted by $y_{2i}^{* (k)}$, from the Bernoulli distribution with the probability of success being equal to $p(x_i, y_{i1}^{* (k)}; \theta_2^{* (k)})$.

[Step 1] **Posterior step (P-step):** Using the current imputed values, generate $\theta_1^{* (k)} = \left( \hat{\beta}_0^{* (k)}, \hat{\beta}_1^{* (k)}, \sigma_{ee}^{* (k)} \right)$ by

$$\sigma_{ee}^{* (k)} \sim n \hat{\sigma}_{ee}^{* (k)} / \chi^2(n)$$

$$\left( \begin{array}{c} \hat{\beta}_0^{* (k)} \\ \hat{\beta}_1^{* (k)} \end{array} \right) \sim N \left( \left( \begin{array}{c} \hat{\beta}_0^{*} \\ \hat{\beta}_1^{*} \end{array} \right), \left( \frac{n}{\sum_{i=1}^{n} x_i} \sum_{i=1}^{n} x_i - 1 \right) \sigma_{ee}^{* (k)} \right),$$

where $\left( \hat{\beta}_0^{* (k)}, \hat{\beta}_1^{* (k)}, \hat{\sigma}_{ee}^{* (k)} \right)$ satisfies

$$\sum_{i=1}^{n} \left( y_{i1}^{* (k)} - \beta_0^{* (k)} - \beta_1^{* (k)} x_i \right) (1, x_i)' = (0, 0)'$$

and

$$\sum_{i=1}^{n} \left( y_{i1}^{* (k)} - \beta_0^{* (k)} - \beta_1^{* (k)} x_i \right)^2 - \sigma_{ee}^{* (k)} = 0.$$

Also, generate $\theta_2^{* (k)} = \left( \hat{\phi}_0^{* (k)}, \hat{\phi}_1^{* (k)}, \hat{\phi}_2^{* (k)} \right)'$ from a normal distribution with mean $\left( \hat{\phi}_0^{* (k)}, \hat{\phi}_1^{* (k)}, \hat{\phi}_2^{* (k)} \right)'$ and $\left( \sum_{i=1}^{n} S_i(\hat{\theta}_2) S_i(\hat{\theta}_2)' \right)^{-1}$, where $\hat{\theta}_2^{* (k)} = (\hat{\phi}_0^{* (k)}, \hat{\phi}_1^{* (k)}, \hat{\phi}_2^{* (k)})$ satisfies

$$\sum_{i=1}^{n} S_i^{* (k)}(\theta_2) = (0, 0, 0)$$

with

$$S_i^{* (k)}(\theta_2) = \left\{ y_{2i}^{* (k)} - p(x_i, y_{i1}^{* (k)}; \theta_2) \right\} \left( 1, x_i, y_{i1}^{* (k)} \right)$$

6
[Step 2] **Imputation step (I-step):** For $\delta_{1i} = 0$, generate the $k$-th imputed values of $y_{1i}$, denoted by $y_{1i}^{* (k)}$, from $N(\beta_0^{(k)} + \beta_1^{(k)} x_i, \sigma_{ee}^{(k)})$. For $\delta_{1i} = 1$, simply set $y_{1i}^{* (k)} = y_{1i}$.

If $\delta_{2i} = 0$, generate the $k$-th imputed values of $y_{2i}$, denoted by $y_{2i}^{* (k)}$, from the Bernoulli distribution with the probability of success being equal to $p(x_i, y_{1i}^{* (k)}, \theta_2^{* (k)})$. For $\delta_{2i} = 1$, simply set $y_{2i}^{* (k)} = y_{2i}$.

[Step 3] Goto Step 1 until $(\theta_1^{* (k)}, \theta_2^{* (k)})$ reaches to a stable distribution.
(Here, simply iterate 100 times.)

Once the multiple imputation is created, the point estimator of $\eta_g = E \{g(Y_1, Y_2)\}$ can be computed by

$$\hat{\eta}_{g, MI} = \frac{1}{M} \sum_{k=1}^{M} \hat{\eta}_g^{(k)}$$

where

$$\hat{\eta}_g^{(k)} = \frac{1}{n} \sum_{i=1}^{n} g\left(y_{1i}^{* (k)}, y_{2i}^{* (k)}\right) := \frac{1}{n} \sum_{i=1}^{n} g_i^{(k)}.$$  

Variance estimation of $\hat{\eta}_{g, MI}$ can be computed by

$$\hat{V}_{MI}(\hat{\eta}_{g, MI}) = \frac{1}{M} \sum_{k=1}^{M} \hat{V}(\hat{\eta}_g^{(k)}) + \left(1 + \frac{1}{M}\right) \frac{1}{M-1} \sum_{k=1}^{M} (\hat{\eta}_g^{(k)} - \hat{\eta}_{g, MI})^2$$

where

$$\hat{V}(\hat{\eta}_g^{(k)}) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(g_i^{(k)} - \hat{\eta}_g^{(k)}\right)^2.$$  

### 4 Simulation Results

Monte Carlo samples of size $n = 200$ were generated independently $B = 2,000$ times. Monte Carlo biases and variances were computed for each
Table 1 Monte Carlo bias and variance of the point estimators.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimator</th>
<th>Bias</th>
<th>Variance</th>
<th>Std Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta_1 )</td>
<td>Complete sample</td>
<td>0.00</td>
<td>.00739</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FI (( M = 100 ))</td>
<td>0.00</td>
<td>.00957</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 100 ))</td>
<td>0.00</td>
<td>.00952</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>CFI (( M = 10 ))</td>
<td>0.00</td>
<td>.00957</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 10 ))</td>
<td>0.00</td>
<td>.00973</td>
<td>132</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>Complete sample</td>
<td>0.00</td>
<td>.00123</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FI (( M = 100 ))</td>
<td>0.00</td>
<td>.00169</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 100 ))</td>
<td>0.00</td>
<td>.00167</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>CFI (( M = 10 ))</td>
<td>0.00</td>
<td>.00169</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 10 ))</td>
<td>0.00</td>
<td>.00170</td>
<td>138</td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>Complete sample</td>
<td>0.00</td>
<td>.00506</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FI (( M = 100 ))</td>
<td>0.00</td>
<td>.00761</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 100 ))</td>
<td>0.00</td>
<td>.00761</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>CFI (( M = 10 ))</td>
<td>0.00</td>
<td>.00761</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 10 ))</td>
<td>0.00</td>
<td>.00788</td>
<td>156</td>
</tr>
<tr>
<td>( \eta_4 )</td>
<td>Complete sample</td>
<td>0.00</td>
<td>.00108</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FI (( M = 100 ))</td>
<td>0.00</td>
<td>.00119</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 100 ))</td>
<td>0.00</td>
<td>.00119</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>CFI (( M = 10 ))</td>
<td>0.00</td>
<td>.00119</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>MI (( M = 10 ))</td>
<td>0.00</td>
<td>.00120</td>
<td>111</td>
</tr>
</tbody>
</table>

estimators and are reported in Table 1. Also, in Table 2, Monte Carlo biases of the variance estimators are reported.
Table 2  Monte Carlo relative bias of the variance estimator.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Imputation</th>
<th>Relative bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V(\hat{\eta}_1)$</td>
<td>FI ($M = 100$)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 100$)</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>CFI ($M = 10$)</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 10$)</td>
<td>0.4</td>
</tr>
<tr>
<td>$V(\hat{\eta}_2)$</td>
<td>FI ($M = 100$)</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 100$)</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>CFI ($M = 10$)</td>
<td>-2.1</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 10$)</td>
<td>0.1</td>
</tr>
<tr>
<td>$V(\hat{\eta}_3)$</td>
<td>FI ($M = 100$)</td>
<td>-2.6</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 100$)</td>
<td>-1.4</td>
</tr>
<tr>
<td></td>
<td>CFI ($M = 10$)</td>
<td>-2.9</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 10$)</td>
<td>-2.3</td>
</tr>
<tr>
<td>$V(\hat{\eta}_4)$</td>
<td>FI ($M = 100$)</td>
<td>-3.2</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 100$)</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>CFI ($M = 10$)</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>MI ($M = 10$)</td>
<td>12.7</td>
</tr>
</tbody>
</table>