A note on multiple imputation for general purpose estimation

Shu Yang    Jae Kwang Kim

SSC meeting

June 16, 2015
Assume simple random sampling, for simplicity.

Under complete response, suppose that

$$\hat{\eta}_{n,g} = n^{-1} \sum_{i=1}^{n} g(y_i)$$

is an unbiased estimator of $\eta_g = E\{g(Y)\}$, for known $g(\cdot)$.

$\delta_i = 1$ if $y_i$ is observed and $\delta_i = 0$ otherwise.

$y_i^*$: imputed value for $y_i$ for unit $i$ with $\delta_i = 0$.

Imputed estimator of $\eta_g$

$$\hat{\eta}_{I,g} = n^{-1} \sum_{i=1}^{n} \{\delta_i g(y_i) + (1 - \delta_i) g(y_i^*)\}$$

Need $E\{g(y_i^*) \mid \delta_i = 0\} = E\{g(y_i) \mid \delta_i = 0\}$. 
Often, find $x$ (always observed) such that

- Missing at random (MAR) holds: $f(y \mid x, \delta = 0) = f(y \mid x)$
- Imputed values are created from $f(y \mid x)$.

Computing the conditional expectation can be a challenging problem.

1. Do not know the true parameter $\theta$ in $f(y \mid x) = f(y \mid x; \theta)$:

$$E \{g(y) \mid x\} = E \{g(y) \mid x; \theta\}.$$

2. Even if we know $\theta$, computing the conditional expectation can be numerically difficult.
Imputation: Monte Carlo approximation of the conditional expectation (given the observed data).

\[
E \{ g(y_i) \mid x_i \} \approx \frac{1}{m} \sum_{j=1}^{m} g\left( y_i^{*}(j) \right)
\]

1. Bayesian approach: generate \( y_i^* \) from

\[
f(y_i \mid x_i, y_{obs}) = \int f(y_i \mid x_i, \theta) p(\theta \mid x_i, y_{obs}) d\theta
\]

2. Frequentist approach: generate \( y_i^* \) from \( f\left( y_i \mid x_i; \hat{\theta} \right) \), where \( \hat{\theta} \) is a consistent estimator.
Thus, imputation is a computational tool for computing the conditional expectation $E\{g(y_i) \mid x_i\}$ for missing unit $i$.

To compute the conditional expectation, we need to specify a model $f(y \mid x; \theta)$ evaluated at $\theta = \hat{\theta}$.

Thus, we can write $\hat{\eta}_{l,g} = \hat{\eta}_{l,g}(\hat{\theta})$.

To estimate the variance of $\hat{\eta}_{l,g}$, we need to take into account of the sampling variability of $\hat{\theta}$ in $\hat{\eta}_{l,g} = \hat{\theta}_{l,g}(\hat{\theta})$. 
Three approaches

- **Bayesian approach**: multiple imputation by Rubin (1978, 1987), Rubin and Schenker (1986), etc.
## Comparison

<table>
<thead>
<tr>
<th></th>
<th>Bayesian</th>
<th>Frequentist</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Posterior distribution $f(\text{latent, } \theta \mid \text{data})$</td>
<td>Prediction model $f(\text{latent } \mid \text{data, } \theta)$</td>
</tr>
<tr>
<td><strong>Computation</strong></td>
<td>Data augmentation</td>
<td>EM algorithm</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td>I-step</td>
<td>E-step</td>
</tr>
<tr>
<td><strong>Parameter update</strong></td>
<td>P-step</td>
<td>M-step</td>
</tr>
<tr>
<td><strong>Parameter est’n</strong></td>
<td>Posterior mode</td>
<td>ML estimation</td>
</tr>
<tr>
<td><strong>Imputation</strong></td>
<td>Multiple imputation</td>
<td>Fractional imputation</td>
</tr>
<tr>
<td><strong>Variance estimation</strong></td>
<td>Rubin’s formula</td>
<td>Linearization or Bootstrap</td>
</tr>
</tbody>
</table>
Step 1. (Imputation) Create $M$ complete datasets by filling in missing values with imputed values generated from the posterior predictive distribution.

To create the $j$th imputed dataset, first generate $\theta^*(j)$ from the posterior distribution $p(\theta \mid X_n, y_{obs})$, and then generate $y^*_i(j)$ from the imputation model $f(y \mid x_i; \theta^*(j))$ for each missing $y_i$.

Step 2. (Analysis) Apply the user’s complete-sample estimation procedure to each imputed dataset.

Let $\hat{\eta}(j)$ be the complete-sample estimator of $\eta = E\{g(Y)\}$ applied to the $j$th imputed dataset and $\hat{\nu}(j)$ be the complete-sample variance estimator of $\hat{\eta}(j)$. 
Step 3. (Summarize) Use Rubin’s combining rule to summarize the results from the multiply imputed datasets.

- The multiple imputation estimator of $\eta$, denoted by $\hat{\eta}_{MI}$, is

$$\hat{\eta}_{MI} = \frac{1}{M} \sum_{j=1}^{M} \hat{\eta}_i^{(j)}$$

- Rubin’s variance estimator is

$$\hat{V}_{MI}(\hat{\eta}_{MI}) = W_M + \left(1 + \frac{1}{M}\right) B_M,$$

where $W_M = M^{-1} \sum_{j=1}^{M} \hat{V}^{(j)}$ and $B_M = (M - 1)^{-1} \sum_{j=1}^{M} (\hat{\eta}_i^{(j)} - \hat{\eta}_{MI})^2$. 
Motivating Example

Suppose that you are interested in estimating $\eta = P(Y \leq 3)$.

Assume a normal model for $f(y | x; \theta)$ for multiple imputation.

Two choices for $\hat{\eta}_n$:

1. Method-of-moments (MOM) estimator: $\hat{\eta}_{n1} = n^{-1} \sum_{i=1}^{n} I(y_i \leq 3)$.

2. Maximum-likelihood estimator:

$$\hat{\eta}_{n2} = n^{-1} \sum_{i=1}^{n} P(Y \leq 3 | x_i; \hat{\theta}),$$

where $P(Y \leq 3 | x_i; \hat{\theta}) = \int_{-\infty}^{3} f(y | x_i; \hat{\theta}) dy$.

Rubin’s variance estimator is nearly unbiased for $\hat{\eta}_{n2}$, but provide conservative variance estimation for $\hat{\eta}_{n1}$ (30-50% overestimation of the variances in most cases).
The goal is to

1. understand why MI provide biased variance estimation for MOM estimation.

2. characterize the asymptotic bias of MI variance estimator when MOM estimator is used in the complete-sample analysis;

3. give an alternative variance estimator that can provide asymptotically valid inference for MOM estimation.
Rubin’s variance estimator is based on the following decomposition,

$$\text{var}(\hat{\eta}_{MI}) = \text{var}(\hat{\eta}_n) + \text{var}(\hat{\eta}_{MI,\infty} - \hat{\eta}_n) + \text{var}(\hat{\eta}_{MI} - \hat{\eta}_{MI,\infty}),$$  \hspace{1cm} (1)

where \(\hat{\eta}_n\) is the complete-sample estimator of \(\eta\) and \(\hat{\eta}_{MI,\infty}\) is the probability limit of \(\hat{\eta}_{MI}\) for \(M \to \infty\).

Under some regularity conditions, \(W_M\) term estimates the first term, the \(B_M\) term estimates the second term, and the \(M^{-1}B_M\) term estimates the last term of (1), respectively.
In particular, Kim et al (2006, JRSSB) proved that the bias of Rubin’s variance estimator is

$$\text{Bias}(\hat{V}_{MI}) \approx -2 \text{cov}(\hat{\eta}_{MI} - \hat{\eta}_n, \hat{\eta}_n).$$  \hspace{1cm} (2)

The decomposition (1) is equivalent to assuming that

$$\text{cov}(\hat{\eta}_{MI} - \hat{\eta}_n, \hat{\eta}_n) \approx 0,$$

which is called the congeniality condition by Meng (1994).

The congeniality condition holds when $\hat{\eta}_n$ is the MLE of $\eta$. In such cases, Rubin’s variance estimator is asymptotically unbiased.

If method of moment (MOM) estimator is used to estimate $\eta = E\{g(Y)\}$, Rubin’s variance estimator can be asymptotically biased.
Assume that \( y \) is observed for the first \( r \) elements.

MI for MOM estimator of \( \eta = E\{g(Y)\} \):

\[
\hat{\eta}_MI = \frac{1}{n} \sum_{i=1}^{r} g(y_i) + \frac{1}{n \cdot M} \sum_{i=r+1}^{n} \sum_{j=1}^{M} g(y_{i}^{*(j)})
\]

where \( y_{i}^{*(j)} \) are generated from \( f(y \mid \theta^{*(j)}) \) and \( \theta^{*(j)} \) are generated from \( p(\theta \mid y_{obs}) \).
Since, conditional on the observed data

\[ p \lim_{M \to \infty} \frac{1}{M} \sum_{j=1}^{M} g(y_i^{*(j)}) = E \left[ E \{ g(Y); \theta^* \} \mid y_{obs} \right] \approx E \{ g(Y); \hat{\theta} \} \]

and the MI estimator of \( \eta \) (for \( M \to \infty \)) can be written

\[ \hat{\eta}_{MI} = \frac{r}{n} \hat{\eta}_{MME,r} + \frac{n-r}{n} \hat{\eta}_{MLE,r}. \]

Thus, \( \hat{\eta}_{MI} \) is a convex combination of \( \hat{\eta}_{MME,r} \) and \( \hat{\eta}_{MLE,r} \), where

\( \hat{\eta}_{MLE,r} = E \{ g(Y); \hat{\theta} \} \) and \( \hat{\eta}_{MME,r} = r^{-1} \sum_{i=1}^{r} g(y_i) \).
Since

\[ \text{Bias}(\hat{V}_{\text{MI}}) \cong -2 \text{Cov}(\hat{\eta}_{\text{MME},n}, \hat{\eta}_{\text{MI}} - \hat{\eta}_{\text{MME},n}), \]

we have only to evaluate the covariance term.

Writing

\[ \hat{\eta}_{\text{MME},n} = p \cdot \hat{\eta}_{\text{MME},r} + (1 - p) \cdot \hat{\eta}_{\text{MME},n-r}, \]

where \( p = r/n \), we can obtain

\[
\text{Cov}(\hat{\eta}_{\text{MME},n}, \hat{\eta}_{\text{MI}} - \hat{\eta}_{\text{MME},n}) \\
= \text{Cov}\{\hat{\eta}_{\text{MME},n}, (1 - p)(\hat{\eta}_{\text{MLE},r} - \hat{\eta}_{\text{MME},n-r})\} \\
= p(1 - p) \text{Cov}\{\hat{\eta}_{\text{MME},r}, \hat{\eta}_{\text{MLE},r}\} - (1 - p)^2 \text{V}\{\hat{\eta}_{\text{MME},n-r}\} \\
= p(1 - p) \{ \text{V}(\hat{\eta}_{\text{MLE},r}) - \text{V}(\hat{\eta}_{\text{MME},r}) \},
\]

which proves \( \text{Bias}(\hat{V}_{\text{MI}}) \geq 0 \).
Bias of Rubin’s variance estimator

Theorem

Let \( \hat{\eta}_n = n^{-1} \sum_{i=1}^{n} g(y_i) \) be the method of moments estimator of \( \eta = E\{g(Y)\} \) under complete response. Assume that \( E(\hat{V}^{(j)}) \approx \text{var}(\hat{\eta}_n) \) holds for \( j = 1, \ldots, M \). Then for \( M \to \infty \), the bias of Rubin’s variance estimator is

\[
\text{Bias}(\hat{V}_{\text{MI}}) \approx 2n^{-1}(1 - p) \left( E \left[ \text{var}\{g(Y) \mid X\} \mid \delta = 0 \right] - \hat{m}_{\theta,0}^T I_{\theta}^{-1} \hat{m}_{\theta,1} \right),
\]

where \( p = r/n \), \( I_{\theta} = -E\{\partial^2 \log f(Y \mid X; \theta) / \partial \theta \partial \theta^T\} \),

\[
m(x; \theta) = E\{g(Y) \mid x; \theta\}, \quad \hat{m}_{\theta}(x) = \partial m(x; \theta) / \partial \theta,
\]

\[
\hat{m}_{\theta,0} = E\{\hat{m}_{\theta}(X) \mid \delta = 0\}, \quad \text{and} \quad \hat{m}_{\theta,1} = E\{\hat{m}_{\theta}(X) \mid \delta = 1\}.
\]
Under MCAR, the bias simplifies to

$$\text{Bias}(\hat{V}_{MI}) \approx 2p(1 - p)\{\text{var}(\hat{\eta}_r,\text{MME}) - \text{var}(\hat{\eta}_r,\text{MLE})\},$$

where $\hat{\eta}_r,\text{MME} = r^{-1} \sum_{i=1}^{r} g(y_i)$ and $\hat{\eta}_r,\text{MLE} = r^{-1} \sum_{i=1}^{r} E\{g(Y) \mid x_i; \hat{\theta}\}$.

This shows that Rubin’s variance estimator is unbiased if and only if the method of moments estimator is as efficient as the maximum likelihood estimator, that is, $\text{var}(\hat{\eta}_r,\text{MME}) \approx \text{var}(\hat{\eta}_r,\text{MLE})$.

Otherwise, Rubin’s variance estimator is positively biased.
Rubin’s variance estimator can be negatively biased

- Now consider a simple linear regression model which contains one covariate $X$ and no intercept.
- By Theorem 1,

$$\text{Bias}(\hat{V}_{MI}) \approx \frac{2(1 - p)\sigma^2}{n} \left\{ 1 - \frac{E(X | \delta = 0)E(X | \delta = 1)}{E(X^2 | \delta = 1)} \right\},$$

which can be zero, positive or negative.
- If

$$E(X | \delta = 0) > \frac{E(X^2 | \delta = 1)}{E(X | \delta = 1)}$$

the Rubins’ variance estimator can be negatively biased.
Decompose

\[
\text{var} \left( \hat{\eta}_{\text{MI}, \infty} \right) = n^{-1} V_1 + r^{-1} V_2,
\]

where

\[
V_1 = \text{var}\{g(Y)\} - (1 - p) E[\text{var}\{g(Y) \mid X\} \mid \delta = 0],
\]

and

\[
V_2 = \dot{m}_\theta^T I_\theta^{-1} \dot{m}_\theta - p^2 \dot{m}_{\theta,1}^T I_\theta^{-1} \dot{m}_{\theta,1}.
\]

The first term \( n^{-1} V_1 \), is the variance of the sample mean of \( \delta_i g(y_i) + (1 - \delta_i) m(x_i; \theta) \).

The second term \( r^{-1} V_2 \), reflects the variability associated with the estimated value of \( \theta \) instead of the true value \( \theta \) in the imputed values.
The first term $n^{-1}V_1$ can be estimated by $\tilde{W}_M = W_M - C_M$, where

\begin{align*}
W_M &= M^{-1} \sum_{j=1}^{M} \hat{V}(j), \\
C_M &= \frac{1}{n^2(M-1)} \sum_{k=1}^{M} \sum_{i=r+1}^{n} \left\{ g(y_i^{*(k)}) - \frac{1}{M} \sum_{k=1}^{M} g(y_i^{*(k)}) \right\}^2.
\end{align*}

since $E\{ W_M \} \approx n^{-1} \text{var}\{ g(Y) \}$ and $E(C_M) \approx n^{-1}(1 - p)E[\text{var}\{ g(Y) \mid X \} \mid \delta = 0]$. 
To estimate the second term term, we use over-imputation:

Table: Over-Imputation Data Structure

<table>
<thead>
<tr>
<th>Row</th>
<th>1</th>
<th>…</th>
<th>M</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(g_1^{(1)})</td>
<td>…</td>
<td>(g_1^{(M)})</td>
<td>(\bar{g}_1^*)</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>(r)</td>
<td>(g_r^{(1)})</td>
<td>…</td>
<td>(g_r^{(M)})</td>
<td>(\bar{g}_r^*)</td>
</tr>
<tr>
<td>(r+1)</td>
<td>(g_{r+1}^{(1)})</td>
<td>…</td>
<td>(g_{r+1}^{(M)})</td>
<td>(\bar{g}_{r+1}^*)</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>(n)</td>
<td>(g_n^{(1)})</td>
<td>…</td>
<td>(g_n^{(M)})</td>
<td>(\bar{g}_n^*)</td>
</tr>
<tr>
<td>average</td>
<td>(\bar{\eta}_n^{(1)})</td>
<td>…</td>
<td>(\bar{\eta}_n^{(M)})</td>
<td>(\bar{\eta}_n^*)</td>
</tr>
</tbody>
</table>
The second term $r^{-1}V_2$ can be estimated by $D_M = D_{M,n} - D_{M,r}$, where

$$D_{M,n} = (M-1)^{-1} \sum_{k=1}^{M} (n^{-1} \sum_{i=1}^{n} d_i^{*(k)})^2 - (M-1)^{-1} \sum_{k=1}^{M} n^{-2} \sum_{i=1}^{n} (d_i^{*(k)})^2,$$

$$D_{M,r} = (M-1)^{-1} \sum_{k=1}^{M} (n^{-1} \sum_{i=1}^{r} d_i^{*(k)})^2 - (M-1)^{-1} \sum_{k=1}^{M} n^{-2} \sum_{i=1}^{r} (d_i^{*(k)})^2,$$

with $d_i^{(k)} = g(y_i^{(k)}) - M^{-1} \sum_{l=1}^{M} g(y_i^{(l)})$.

since $E(D_{M,n}) \approx r^{-1} \hat{m}_\theta^T \mathcal{I}_\theta^{-1} \hat{m}_\theta$ and $E(D_{M,r}) \approx r^{-1} p^2 \hat{m}_{\theta,1}^T \mathcal{I}_\theta^{-1} \hat{m}_{\theta,1}$. 
Theorem

Under the assumptions of Theorem 1, the new multiple imputation variance estimator is

\[ \hat{V}_{MI} = \tilde{W}_M + D_M + M^{-1}B_M, \]

where \( \tilde{W}_M = W_M - C_M \), and \( B_M \) being the usual between-imputation variance. \( \hat{V}_{MI} \) is asymptotically unbiased for estimating the variance of the multiple imputation estimator as \( n \to \infty \).
Samples of size $n = 2,000$ are independently generated from

$$Y_i = \beta X_i + e_i,$$

where $\beta = 0.1$, $X_i \sim \exp(1)$ and $e_i \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon^2 = 0.5$.

Let $\delta_i$ be the response indicator of $y_i$ and $\delta_i \sim \text{Bernoulli}(p_i)$, where

$$p_i = 1/\{1 + \exp(-\phi_0 - \phi_1 x_i)\}.$$

We consider two scenarios:

(i) $(\phi_0, \phi_1) = (-1.5, 2)$ and (ii) $(\phi_0, \phi_1) = (3, -3)$

The parameters of interest are $\eta_1 = E(Y)$ and $\eta_2 = \text{pr}(Y < 0.15)$. 

Simulation study

Result 1

**Table:** Relative biases of two variance estimators and mean width and coverages of two interval estimates under two scenarios in simulation one

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( \eta_1 )</th>
<th>( \eta_2 )</th>
<th>Relative bias (%)</th>
<th>Mean Width for 95% C.I.</th>
<th>Coverage for 95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \eta_1 )</td>
<td>96.8</td>
<td>0.7</td>
<td>0.038</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>( \eta_2 )</td>
<td>123.7</td>
<td>2.9</td>
<td>0.027</td>
<td>0.018</td>
</tr>
<tr>
<td>2</td>
<td>( \eta_1 )</td>
<td>-19.8</td>
<td>0.4</td>
<td>0.061</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>( \eta_2 )</td>
<td>-9.6</td>
<td>-0.4</td>
<td>0.037</td>
<td>0.039</td>
</tr>
</tbody>
</table>

C.I., confidence interval; \( \eta_1 = E(Y) \); \( \eta_2 = \Pr(Y < 0.15) \).
Simulation study

Result 1

- For $\eta_1 = E(Y)$, under scenario (i), the relative bias of Rubin’s variance estimator is 96.8%, which is consistent with our result with $E_1(X^2) - E_0(X)E_1(X) > 0$, where $E_1(X^2) = 3.38$, $E_1(X) = 1.45$, and $E_0(X) = 0.48$.

- Under scenario (ii), the relative bias of Rubin’s variance estimator is $-19.8\%$, which is consistent with our result with $E_1(X^2) - E_0(X)E_1(X) < 0$, where $E_1(X^2) = 0.37$, $E_1(X) = 0.47$, and $E_0(X) = 1.73$.

- The empirical coverage for Rubin’s method can be over or below the nominal coverage due to variance overestimation or underestimation.

- On the other hand, the new variance estimator is essentially unbiased for these scenarios.
Simulation study
Set up 2

- Samples of size \( n = 200 \) are independently generated from

\[
Y_i = \beta_0 + \beta_1 X_i + e_i,
\]

where \( \beta = (\beta_0, \beta_1) = (3, -1) \), \( X_i \sim N(2, 1) \) and \( e_i \sim N(0, \sigma_e^2) \) with \( \sigma_e^2 = 1 \).

- The parameters of interest are \( \eta_1 = E(Y) \) and \( \eta_2 = \text{pr}(Y < 1) \).

- We consider two different factors:
  1. The response mechanism: MCAR and MAR:
     For MCAR, \( \delta_i \sim \text{Bernoulli}(0.6) \).
     For MAR, \( \delta_i \sim \text{Bernoulli}(p_i) \), where \( p_i = 1/\{1 + \exp(-0.28 - 0.1x_i)\} \) with the average response rate about 0.6.
  2. The size of multiple imputation, with two levels \( M = 10 \) and \( M = 30 \).
Table: Relative biases of two variance estimators and mean width and coverages of two interval estimates under two scenarios of missingness in simulation two

<table>
<thead>
<tr>
<th></th>
<th>Relative Bias (%)</th>
<th>Mean Width for 95% C.I.</th>
<th>Coverage for 95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M Rubin New Rubin</td>
<td>New</td>
<td>Rubin New</td>
</tr>
<tr>
<td>Missing completely at random</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>10 -0.9 -1.58</td>
<td>0.240 0.250</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td></td>
<td>30 -0.6 -1.68</td>
<td>0.230 0.235</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>10 22.7 -1.14</td>
<td>0.083 0.083</td>
<td>0.98 0.95</td>
</tr>
<tr>
<td></td>
<td>30 23.8 -1.23</td>
<td>0.082 0.075</td>
<td>0.98 0.95</td>
</tr>
<tr>
<td>Missing at random</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>10 -1.0 -1.48</td>
<td>0.23 0.25</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td></td>
<td>30 -0.9 -1.59</td>
<td>0.231 0.23</td>
<td>0.95 0.95</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>10 20.7 -1.64</td>
<td>0.081 0.081</td>
<td>0.98 0.95</td>
</tr>
<tr>
<td></td>
<td>30 21.5 -1.74</td>
<td>0.074 0.071</td>
<td>0.98 0.95</td>
</tr>
</tbody>
</table>

C.I., confidence interval; \( \eta_1 = E(Y) \); \( \eta_2 = \text{pr}(Y < 1) \).
Both Rubin’s variance estimator and our new variance estimator are unbiased for $\eta_1 = E(Y)$.

Rubin’s variance estimator is biased upward for $\eta_2 = \Pr(Y < 1)$, with absolute relative bias as high as 24%; whereas our new variance estimator reduces absolute relative bias to less than 1.74%.

For Rubin’s method, the empirical coverage for $\eta_2 = \Pr(Y < 1)$ reaches to 98% for 95% confidence intervals, due to variance overestimation.

In contrast, our new method provides more accurate coverage of confidence interval for both $\eta_1 = E(Y)$ and $\eta_2 = \Pr(Y < 1)$ at 95% levels.
Conclusion

- Investigate asymptotic properties of Rubin’s variance estimator.
- If method of moment is used, Rubin’s variance estimator can be asymptotically biased.
- New variance estimator, based on multiple over-imputation, can provide valid variance estimation in this case.
- Our method can be extended to a more general class of parameters obtained from estimating equations.

\[
\sum_{i=1}^{n} U(\eta; x_i, y_i) = 0.
\]

This is a topic of future study.
Thank you!