§3 Variance estimation after imputation

Recall Example 4.2

- Under the setup of Example 4.1, we are interested in estimating the asymptotic variance of the regression imputation estimator

$$\hat{\theta}_{\mathrm{I,reg}} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right), \tag{1}$$

where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)'$ is the solution to

$$\hat{U}(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n}\delta_i(y_i - \beta_0 - \beta_1 x_i)\begin{pmatrix}1\\ x_i\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}.$$

- Using Taylor linearization, we obtain

$$\begin{aligned}\hat{\theta}_{\mathrm{I,reg}} &= \hat{\theta}_{\mathrm{I}}(\boldsymbol{\beta}^*) + (\kappa_1, \kappa_2)\hat{U}(\boldsymbol{\beta}^*) + o_p(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{(\beta_0 + \beta_1 x_i) + \delta_i(y_i - \beta_0 - \beta_1 x_i)(\kappa_1 + \kappa_2 x_i)\right\} + o_p(n^{-1/2}),\end{aligned}$$

where

$$\begin{pmatrix}\kappa_1\\ \kappa_2\end{pmatrix} = \begin{pmatrix}E(\delta) & E(\delta X)\\ E(\delta X) & E(\delta X^2)\end{pmatrix}^{-1}\begin{pmatrix}1\\ E(X)\end{pmatrix}.$$

- Note that

$$d(x_i, y_i, y_i; \beta, \kappa) = (\beta_0 + \beta_1 x_i) + \delta_i (y_i - \beta_0 - \beta_1 x_i)(\kappa_1 + \kappa_2 x_i)$$

  are IID.

- For variance estimation of $\hat{\theta}_{\mathrm{I,reg}}$, we can consider estimating the variance of $\bar{d}_n = n^{-1} \sum_{i=1}^{n} d_i$, where $d_i = d(x_i, y_i, y_i; \beta, \kappa)$. That is, we can use

$$\hat{V} = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{d}_i - \bar{d}_n \right)^2$$

  as a linearized variance estimator, where $\hat{d}_i = d(x_i, y_i, \delta_i; \hat{\beta}, \hat{\kappa})$ and

$$\begin{pmatrix} \hat{\kappa}_1 \\ \hat{\kappa}_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} \delta_i & \sum_{i=1}^{n} \delta_i x_i \\ \sum_{i=1}^{n} \delta_i x_i & \sum_{i=1}^{n} \delta_i x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} n \\ \sum_{i=1}^{n} x_i \end{pmatrix}. \qquad (2)$$

Remark 1

- The regression imputation estimator in (1) is linear in y. Thus, we can express

$$\hat{\theta}_{\mathrm{I,reg}} = \frac{1}{n}\sum_{i=1}^{n} \delta_i w_i y_i \tag{3}$$

  where

$$w_i = \hat{\kappa}_1 + \hat{\kappa}_2 x_i$$

  and $\hat{\kappa} = (\hat{\kappa}_1, \hat{\kappa}_2)'$ is defined in (2).

- The pseudo value $\hat{d}_i$ for variance estimation can be expressed as

$$\hat{d}_i = \hat{y}_i + \delta_i w_i (y_i - \hat{y}_i)$$

  where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

Remark 2

- The weight w_i in (3) satisfies

$$\sum_{i=1}^{n} \delta_i w_i (1, x_i) = \sum_{i=1}^{n} (1, x_i). \tag{4}$$

- Equation (4) is often called calibration equation.
- In fact, w_i in (3) can be obtained by solving the following optimization problem: Minimize

$$\sum_{i=1}^{n} \delta_i w_i^2$$

subject to (4).

Deterministic Imputation

- In Example 4.2, the imputed estimator can be written as $\hat{\theta}_{Id}(\hat{\beta})$. Note that we can write the deterministic imputed estimator as

$$\hat{\theta}_{Id} = n^{-1} \sum_{i=1}^{n} \hat{y}_i(\hat{\beta}),$$

  where $\hat{y}_i(\hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
- In general, the asymptotic variance of $\hat{\theta}_{Id} = \hat{\theta}(\hat{\beta})$ is different from the asymptotic variance of $\hat{\theta}(\beta)$

- As in Example 4.2, if we can find $d_i = d_i(\beta)$ such that

$$\hat{\theta}_{Id}(\hat{\beta}) = n^{-1} \sum_{i=1}^{n} d_i(\hat{\beta}) \cong n^{-1} \sum_{i=1}^{n} d_i(\beta),$$

  then the asymptotic variance of $\hat{\theta}_{Id}$ is equal to the asymptotic variance of $\bar{d}_n = n^{-1} \sum_{i=1}^{n} d_i(\beta)$.
- Note that, if (x_i, y_i, δ_i) are IID, then $d_i = d(x_i, y_i, \delta_i)$ are also IID. Thus, the variance of $\bar{d}_n = n^{-1} \sum_{i=1}^{n} d_i$ is unbiasedly estimated by

$$\hat{V}(\bar{d}_n) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(d_i - \bar{d}_n\right)^2 . \qquad (5)$$

  Unfortunately, we cannot compute $\hat{V}(\bar{d}_n)$ in (5) since $d_i = d_i(\beta)$ is a function of unknown parameters. Thus, we use $\hat{d}_i = d_i(\hat{\beta})$ in (5) to get a consistent variance estimator of the imputed estimator.

Stochastic Imputation

- Instead of the deterministic imputation, suppose that a stochastic imputation is used such that

$$\hat{\theta}_I = n^{-1} \sum_{i=1}^{n} \left\{ \delta_i y_i + (1 - \delta_i) \left(\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{e}_i^* \right) \right\},$$

  where $\hat{e}_i^*$ are the additional noise terms in the stochastic imputation. Often $\hat{e}_i^*$ are randomly selected from the empirical distribution of the sample residuals in the respondents.

- The variance of the imputed estimator can be decomposed into two parts:

$$V\left(\hat{\theta}_I\right) = V\left(\hat{\theta}_{Id}\right) + V\left(\hat{\theta}_I - \hat{\theta}_{Id}\right) \tag{6}$$

  where the first part is the deterministic part and the second part is the additional variance due to stochastic imputation. The first part can be estimated by the linearization method discussed above. The second part is called the imputation variance.

- If we require the imputation mechanism to satisfy

$$\sum_{i=1}^{n} (1 - \delta_i) \hat{e}_i^* = 0$$

then the imputation variance is equal to zero.
- Often the variance of $\hat{\theta}_I - \hat{\theta}_{Id} = n^{-1} \sum_{i=1}^{n} (1 - \delta_i) \hat{e}_i^*$ can be computed under the known imputation mechanism. For example, if simple random sampling without replacement is used then

$$V\left(\hat{\theta}_I - \hat{\theta}_{Id}\right) = \frac{n-r}{n^2}\left(1 - \frac{n-r}{r}\right)\frac{1}{r-1}\sum_{i=1}^{n} \delta_i \hat{e}_i^2.$$

Imputed estimator for general parameters

- We now discuss a general case of parameter estimation when the parameter of interest ψ is estimated by the solution $\hat{\psi}_n$ to

$$\sum_{i=1}^{n} U(\psi; \mathbf{y}_i) = 0 \tag{7}$$

under complete response of $\mathbf{y}_1, \cdots, \mathbf{y}_n$.

- Under the existence of missing data, we can use the imputed estimating equation

$$\bar{U}_m^*(\psi) \equiv m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} U(\psi; \mathbf{y}_i^{*(j)}) = 0, \tag{8}$$

where $\mathbf{y}_i^{*(j)} = (\mathbf{y}_{i,\mathrm{obs}}, \mathbf{y}_{i,mis}^{*(j)})$ and $\mathbf{y}_{i,mis}^{*(j)}$ are randomly generated from the conditional distribution $p(\mathbf{y}_{i,\mathrm{mis}} \mid \mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i; \hat{\eta}_p)$ where $\hat{\eta}_p$ is estimated by solving

$$\hat{U}_p(\eta) \equiv \sum_{i=1}^{n} U_p(\eta; \mathbf{y}_{i,\mathrm{obs}}) = 0. \tag{9}$$

- To apply the linearization method, we first compute the conditional expectation of $U(\psi; \mathbf{y}_i)$ given $(\mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i)$ evaluated at $\hat{\eta}_p$. That is, compute

$$\bar{U}(\psi \mid \hat{\eta}_p) = \sum_{i=1}^{n} \bar{U}_i(\psi \mid \hat{\eta}_p) = \sum_{i=1}^{n} E\left\{U(\psi; \mathbf{y}_i) \mid \mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i; \hat{\eta}_p\right\}. \tag{10}$$

- Let $\hat{\psi}_R$ be the solution to $\bar{U}(\psi \mid \hat{\eta}_p) = 0$. Using the linearization technique, we have

$$\bar{U}(\psi \mid \hat{\eta}_p) \cong \bar{U}(\psi \mid \eta_0) + E\left\{\frac{\partial}{\partial \eta'} \bar{U}(\psi \mid \eta_0)\right\}(\hat{\eta}_p - \eta_0) \tag{11}$$

and

$$0 = \hat{U}_p(\hat{\eta}_p) = \hat{U}_p(\eta_0) + E\left\{\frac{\partial}{\partial \eta'} \hat{U}_p(\eta_0)\right\}(\hat{\eta}_p - \eta_0). \tag{12}$$

- Thus, combining (11) and (12), we have

$$\bar{U}(\psi \mid \hat{\eta}_p) \cong \bar{U}(\psi \mid \eta_0) + \kappa(\psi)\hat{U}_p(\eta_0) \tag{13}$$

where

$$\kappa(\psi) = -E\left\{\frac{\partial}{\partial \eta'}\bar{U}(\psi \mid \eta_0)\right\}\left[E\left\{\frac{\partial}{\partial \eta'}\hat{U}_p(\eta_0)\right\}\right]^{-1}.$$

- Thus, writing

$$\bar{U}_\ell(\psi \mid \eta_0) = \sum_{i=1}^{n}\left\{\bar{U}_i(\psi \mid \eta_0) + \kappa(\psi)\hat{U}_p(\eta_0; \mathbf{y}_{i,\mathrm{obs}})\right\} = \sum_{i=1}^{n} q_i(\psi \mid \eta_0),$$

and $q_i(\psi \mid \eta_0) = \bar{U}_i(\psi \mid \eta_0) + \kappa(\psi)\hat{U}_p(\eta_0; \mathbf{y}_{i,\mathrm{obs}})$, the variance of $\bar{U}(\psi \mid \hat{\eta}_p)$ is asymptotically equal to the variance of $\bar{U}_\ell(\psi \mid \eta_0)$.

- Thus, the sandwich-type variance estimator for $\hat{\psi}_R$ is

$$\hat{V}\left(\hat{\psi}_R\right) = \hat{\tau}_q^{-1}\hat{\Omega}_q\hat{\tau}_q^{-1'} \tag{14}$$

where

$$\begin{aligned} \hat{\tau}_q &= n^{-1}\sum_{i=1}^{n}\dot{q}_i\left(\hat{\psi}_R \mid \hat{\eta}_p\right) \\ \hat{\Omega}_q &= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\left(\hat{q}_i - \bar{q}_n\right)^{\otimes 2}, \end{aligned}$$

$\dot{q}_i(\psi \mid \eta) = \partial q_i(\psi \mid \eta)/\partial\psi$, $\bar{q}_n = n^{-1}\sum_{i=1}^{n}\hat{q}_i$, and $\hat{q}_i = q_i(\hat{\psi}_R \mid \hat{\eta}_p)$.

- Note that

$$\begin{aligned} \hat{\tau}_q &= n^{-1}\sum_{i=1}^{n}\dot{q}_i\left(\hat{\psi}_R \mid \hat{\eta}_p\right) \\ &= n^{-1}\sum_{i=1}^{n}E\left\{\dot{U}(\hat{\psi}_R;\mathbf{y}_i) \mid \mathbf{y}_{i,\mathrm{obs}},\boldsymbol{\delta}_i;\hat{\eta}_p\right\} \end{aligned}$$

because $\hat{\eta}_p$ is the solution to (9).

Example 4.3

- Assume that the original sample is decomposed into G disjoint groups (often called imputation cells) and the sample observations are IID within the same cell. That is,
$$y_i \mid i \in S_g \overset{i.i.d.}{\sim} \left(\mu_g, \sigma_g^2\right) \tag{15}$$
where S_g is the set of sample indices in cell g. Assume that n_g sample elements in cell g and r_g elements are observed in the cell.
- For deterministic imputation, let $\hat{\mu}_g = r_g^{-1} \sum_{i \in S_g} \delta_i y_i$ be the g-th cell mean of y among the respondents. The (deterministically) imputed estimator of $\theta = E(Y)$ is, under MAR,
$$\hat{\theta}_{Id} = n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \left\{\delta_i y_i + (1 - \delta_i)\hat{\mu}_g\right\} = n^{-1} \sum_{g=1}^{G} n_g \hat{\mu}_g . \tag{16}$$

Example 4.3 (Cont'd)

- Using the linearization technique in (13), the imputed estimator can be expressed as

$$\hat{\theta}_{Id} \cong n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \left\{ \mu_g + \frac{n_g}{r_g} \delta_i (y_i - \mu_g) \right\} \tag{17}$$

and the plug-in variance estimator can be expressed as

$$\hat{V}(\hat{\theta}_{Id}) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{d}_i - \bar{d}_n \right)^2 \tag{18}$$

where $\hat{d}_i = \hat{\mu}_g + (n_g / r_g) \delta_i \left(y_i - \hat{\mu}_g \right)$ and $\bar{d}_n = n^{-1} \sum_{i=1}^{n} \hat{d}_i$.

Example 4.3 (Cont'd)

- If a stochastic imputation is used where an imputed value is randomly selected from the set of respondents in the same cell, then we can write

$$\hat{\theta}_{Is} = n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \left\{ \delta_i y_i + (1 - \delta_i) y_i^* \right\}. \tag{19}$$

Writing $\hat{\theta}_{Is} = \hat{\theta}_{Id} + n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} (1 - \delta_i) \left(y_i^* - \hat{\mu}_g \right)$, the variance of the second term can be estimated by $n^{-2} \sum_{g=1}^{G} \sum_{i \in S_g} (1 - \delta_i) \left(y_i^* - \hat{\mu}_g \right)^2$ if the imputed values are generated independently, conditional on the respondents.