That Prasad-Rao Is Robust: Estimation of Mean Squared Prediction Error of Observed Best Predictor under Potential Model Misspecification

Xiaohui Liu[1], Haiqiang Ma[1] and Jiming Jiang[2]

Jiangxi University of Finance and Economics, China[1]

and University of California, Davis, USA[2]

Supplementary Material

Throughout this supplementary material, the paper by Liu, Ma and Jiang (2020), That Prasad-Rao is robust: Estimation of mean squared prediction error of observed best predictor under potential model misspecification, is referred to as LMJ20. All of the notations used below are consistent with those introduced in LMJ20.

S1 Details in the proof of Theorem 1

The following expression can be derived:

$$g_i(\psi, y_i) - \theta_i = r_i(A)(x_i'\beta - y_i) + e_i$$

with $r_i(A) = D_i/(A + D_i)$. It follows that

$$g_i(\psi, y_i) - g_i(\psi_*, y_i) = r_i(A)x_i'(\beta - \beta_*) + \{r_i(A) - r_i(A_*)\}(x_i'\beta_* - y_i).$$

Under assumptions *A1*, *A2*, and using (2.10) of LMJ20, we have

$$
\begin{aligned}
& \mathrm{E}[\{g_i(\psi_*, y_i) - \theta_i\}\{g_i(\psi, y_i) - g_i(\psi_*, y_i)\}] \\
= \; & r_i(A_*)\mathrm{E}[r_i(A)(x_i'\beta_* - y_i)x_i'(\beta - \beta_*) + \{r_i(A) - r_i(A_*)\}(x_i'\beta_* - y_i)^2] \\
& + r_i(A)\mathrm{E}(e_i x_i')(\beta - \beta_*) + \{r_i(A) - r_i(A_*)\}\mathrm{E}\{e_i(x_i'\beta_* - y_i)\} \\
= \; & r_i(A_*)r_i(A)\mathrm{E}\{(x_i'\beta_* - y_i)x_i'\}(\beta - \beta_*) + r_i(A_*)\{r_i(A) - r_i(A_*)\}\mathrm{E}(y_i - x_i'\beta_*)^2 \\
& + \{r_i(A) - r_i(A_*)\}\mathrm{E}\{e_i(x_i'\beta_* - \theta_i - e_i)\} \\
= \; & r_i(A_*)\{r_i(A) - r_i(A_*)\}(A_* + D_i) - \{r_i(A) - r_i(A_*)\}D_i \\
= \; & \{r_i(A) - r_i(A_*)\}\{r_i(A_*)(A_* + D_i) - D_i\} \\
= \; & 0.
\end{aligned}
$$

Decomposition (2.11) of LMJ20 then follows.

S2 Proof of Theorem 2

By Taylor series expansion (see notation above assumption *A3* in LMJ20), we have

$$
\begin{aligned}
0 &= \frac{\partial \hat{Q}}{\partial \psi} \\
&= \left.\frac{\partial Q}{\partial \psi}\right|_{\psi_*} + \left.\frac{\partial^2 Q}{\partial \psi \partial \psi'}\right|_{\psi_*} (\hat{\psi} - \psi) \\
& \quad + \frac{1}{2}\left[(\hat{\psi} - \psi)' \left.\frac{\partial^3 Q}{\partial \psi_s \partial \psi \partial \psi'}\right|_{\psi_*} (\hat{\psi} - \psi)\right]_{1 \le s \le p+1} + O_{\mathrm{P}}(m^{-1/2})
\end{aligned}
$$

$$= \xi + G(\hat{\psi} - \psi) + \eta(\hat{\psi} - \psi) + \frac{1}{2}[\cdots]_{1\le s\le p+1} + O_{\rm P}(m^{-1/2}), \tag{S2.1}$$

where $\xi = \partial Q/\partial\psi|_{\psi_*}$, $\eta = \partial^2 Q/\partial\psi\partial\psi'|_{\psi_*} - G$ with $G = {\rm E}(\partial^2 Q/\partial\psi\partial\psi'|_{\psi_*})$. From (S2.1), we have $G(\hat{\psi}-\psi) = -\partial Q/\partial\psi + O_{\rm P}(1)$; hence, we have a first-step expansion:

$$\hat{\psi} - \psi = -G^{-1}\xi + O_{\rm P}(m^{-1}). \tag{S2.2}$$

Bringing (S2.2) back to (S2.1), we have

$$0 = \xi + G(\hat{\psi} - \psi) - \eta G^{-1}\xi + \frac{1}{2}[\xi' G^{-1} H_s G^{-1}\xi]_{1\le s\le p+1} + O_{\rm P}(m^{-1/2}),$$

where $H_s = {\rm E}(\partial^3 Q/\partial\psi_s\partial\psi\partial\psi'|_{\psi_*})$. We thus obtain a second-step expansion:

$$\hat{\psi} - \psi = -G^{-1}\xi + G^{-1}\eta G^{-1}\xi - \frac{1}{2}G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\le s\le p+1} + O_{\rm P}(m^{-3/2}). \tag{S2.3}$$

It can be seen that, under the assumptions, there is a constant $\delta > 0$ such that the supremum over $\{\psi : |\psi - \psi_*| \le \delta\}$ of absolute value any partial derivative, up to the 4th order, of Q_j with respect to ψ, $1 \le j \le m$ are bounded. Furthermore, it is seen, by *A3*, that the second moments of any second-order partial derivative of Q_j at ψ_*, with respect to ψ, $1 \le j \le m$ are bounded. Thus, together with the independence assumption (A1 of LMJ20), (S2.1) and (S2.2) are justified. By a similar argument, (S2.3) is justified.

Next, let $\tilde{\theta}_{i*}$ denote $\hat{\theta}_i$ with $\hat{\psi}$ replaced by ψ_*. Then, we have $\hat{\theta}_i - \theta_i =$

$\tilde{\theta}_{i*} - \theta_i + \hat{\theta}_i - \tilde{\theta}_{i*}$; thus, by (2.14) of LMJ20, we have

$$\begin{aligned} \mathrm{MSPE}(\hat{\theta}_i) &= \mathrm{E}(\hat{\theta}_i - \theta_i)^2 \\ &= \frac{A_* D_i}{A_* + D_i} + 2\mathrm{E}\{(\tilde{\theta}_{i*} - \theta_i)(\hat{\theta}_i - \tilde{\theta}_{i*})\} + \mathrm{E}(\hat{\theta}_i - \tilde{\theta}_{i*})^2 \\ &= c_i(A_*) + 2I_1 + I_2, \end{aligned} \tag{S2.4}$$

where I_1 and I_2 are defined in obvious ways.

Let us first consider I_2. We first show that the following holds:

$$\mathrm{E}(\hat{\theta}_i - \tilde{\theta}_{i*})^2 = \mathrm{E}\left(\frac{\partial g_i^*}{\partial \psi'} G^{-1} \xi\right)^2 + o(m^{-1}), \tag{S2.5}$$

where $\partial g_i^* / \partial \psi = \partial g_i / \partial \psi|_{\psi = \psi_*}$. Recall the definition of $\mathcal{A}_\delta$ above *A3* of LMJ20; also define

$$\begin{aligned} \mathcal{B}_\delta &= \left\{ \max_{1 \le s \le p+1} \sup_{|\psi - \psi_*| \le \delta} \left\| \frac{\partial^3 Q}{\partial \psi_s \partial \psi \partial \psi'} \right\| \le c_1 m \right\}, \\ \mathcal{C}_\delta &= \left\{ \max_{1 \le k, l \le p+1} \left| \sum_{j=1}^m \left[\left. \frac{\partial^2 Q_j}{\partial \psi_k \partial \psi_l} \right|_{\psi_*} - \mathrm{E}\left(\left. \frac{\partial^2 Q_j}{\partial \psi_k \partial \psi_l} \right|_{\psi_*} \right) \right] \right| \le \delta_1 m \right\}, \end{aligned}$$

where the spectral norm of a matrix, M, is defined as $\|M\| = \sqrt{\lambda_{\max}(M'M)\}}$ ($\lambda_{\max}$ means largest eigenvalue), and δ, c_1, δ_1 are constants to be determined later. We have

$$\mathrm{E}(\hat{\theta}_i - \tilde{\theta}_{i*})^2 = \mathrm{E}\{(\hat{\theta}_i - \tilde{\theta}_{i*})^2 1_{\mathcal{A}_\delta \cap \mathcal{B}_\delta \cap \mathcal{C}_\delta}\} + \mathrm{E}\{(\hat{\theta}_i - \tilde{\theta}_{i*})^2 1_{\mathcal{A}_\delta^c \cup \mathcal{B}_\delta^c \cup \mathcal{C}_\delta^c}\}. \tag{S2.6}$$

Let c, γ denote generic, positive constants, whose values may be different at different places. We have, by assumption *A4*, that $\mathrm{P}(\mathcal{A}_\delta^c) \le c_\delta m^{-d}$. Also, we have

$$\sup_{|\psi-\psi_*|\le\delta}\left\|\frac{\partial^3 Q}{\partial\psi_s\partial\psi\partial\psi'}\right\| \le \sum_{j=1}^m \sup_{|\psi-\psi_*|\le\delta}\left\|\frac{\partial^3 Q_j}{\partial\psi_s\partial\psi\partial\psi'}\right\| \le c\sum_{j=1}^m (y_j^2+|x_j|^2),$$

if δ is chosen sufficiently small. Write $Z_j = y_j^2 + |x_j|^2$ and $\zeta_j = Z_j - \mathrm{E}(Z_j)$. We have

$$\begin{aligned} \mathrm{P}\left(\sup_{|\psi-\psi_*|\le\delta}\left\|\frac{\partial^3 Q}{\partial\psi_s\partial\psi\partial\psi'}\right\| > c_1 m\right) &\le \mathrm{P}\left(\frac{1}{m}\sum_{j=1}^m Z_j > c\right) \\ &\le \mathrm{P}\left(\frac{1}{m}\sum_{j=1}^m \zeta_j > c - c_2\right) \\ &\le cm^{-2d}\mathrm{E}\left(\left|\sum_{j=1}^m \zeta_j\right|^{2d}\right), \end{aligned}$$

if c_1 is chosen sufficiently large, where c_2 is an upper bound of $\mathrm{E}(Z_j)$. By Marcinkiewicz-Zygmund inequality (e.g., Jiang 2010, p. 150), we have

$$\mathrm{E}\left(\left|\sum_{j=1}^m \zeta_j\right|^{2d}\right) \le c\mathrm{E}\left(\sum_{j=1}^m \zeta_j^2\right)^d \le cm^d\mathrm{E}\left(\frac{1}{m}\sum_{j=1}^m |\zeta_j|^{2d}\right) \le cm^d,$$

using Jensen's inequality for the second-to-last step. It follows that $\mathrm{P}(\mathcal{B}_\delta^c) = O(m^{-d})$. Similarly, it can be shown that $\mathrm{P}(\mathcal{C}_\delta^c) = O(m^{-d})$.

Note that $|\hat{\theta}_i - \theta_{i*}| \le c\{\log(m+1)\}^K(|y_i| + |x_i|)$ by the regularization of $\hat{\psi}$ (above *A3* of LMJ20). Thus, by the Cauchy-Schwarz inequality, we have

$$\mathrm{E}\{(\hat{\theta}_i - \tilde{\theta}_{i*})^2 1_{\mathcal{A}_\delta^c\cup\mathcal{B}_\delta^c\cup\mathcal{C}_\delta^c}\} \le \left[\mathrm{E}\left\{(\hat{\theta}_i - \tilde{\theta}_{i*})^4\right\}\right]^{1/2} \mathrm{P}(\mathcal{A}_\delta^c\cup\mathcal{B}_\delta^c\cup\mathcal{C}_\delta^c)^{1/2} = o(m^{-1}). \tag{S2.7}$$

Now let us see what happens on $\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta$. By Taylor series expansion, we have

$$0 = \xi + \left\{ G + \eta + \frac{1}{2}\left[(\hat{\psi} - \psi_*)' \left. \frac{\partial^3 Q}{\partial \psi_s \partial \psi \partial \psi'} \right|_{\psi_{[s]}} \right]_{1 \le s \le p+1} \right\} (\hat{\psi} - \psi_*), \quad \text{(S2.8)}$$

where $\psi_{[s]}$ lies between $\hat{\psi}$ and ψ_*. It can be shown that

$$G = \sum_{j=1}^{m} \mathrm{E}\left(\left. \frac{\partial^2 Q_j}{\partial \psi \partial \psi'} \right|_{\psi_*} \right) = 2 \begin{bmatrix} s_{0m}\mathrm{E}(x_1 x_1') & 0 \\ 0 & s_{1m} \end{bmatrix},$$

where $s_{km} = \sum_{j=1}^{m} r_j^2/(A_* + D_j)^k$, $k = 0, 1, \ldots$ with $r_j = 1 - B_j = D_j/(A_* + D_j)$. It follows, by assumption *A3*, that there is a positive constant λ such that $G \ge \lambda m I_{p+1}$. Also, by the definition of $\mathcal{A}_\delta$, $\mathcal{B}_\delta$ and $\mathcal{C}_\delta$, it can be seen that

$$\|\eta\| \le (p+1)\delta_1 m, \quad \left\| \left[(\hat{\psi} - \psi_*)' \left. \frac{\partial^3 Q}{\partial \psi_s \partial \psi \partial \psi'} \right|_{\psi_{[s]}} \right]_{1 \le s \le p+1} \right\| \le \delta c m.$$

Thus, by an inequality regarding the smallest and largest eigenvalues of matrices (e.g., Jiang 2010, Exercise 5.27), it can be shown that, by choosing δ, δ_1 sufficiently small, we have $\lambda_{\min}(\{\cdots\}) \ge (\lambda/2)m$, where $\{\cdots\}$ denotes the matrix that $\hat{\psi} - \psi_*$ is multiplied by in (S2.8). It follows that $\{\cdots\}$ is invertible, thus $\hat{\psi} - \psi_* = -\{\cdots\}^{-1}\xi$; hence, we have

$$|\hat{\psi} - \psi_*| \le \|\{\cdots\}^{-1}\| \cdot |\xi| = \frac{|\xi|}{\lambda_{\min}(\{\cdots\})} \le \frac{2}{\lambda}\left(\frac{|\xi|}{m} \right). \quad \text{(S2.9)}$$

Now, again by Taylor series expansion, we have

$$\begin{aligned} \hat{\theta}_i - \tilde{\theta}_{i*} &= g_i(\hat{\psi}, y_i) - g_i(\psi_*, y_i) \\ &= \left. \frac{\partial g_i}{\partial \psi'} \right|_{\psi_*} (\hat{\psi} - \psi_*) + \frac{1}{2}(\hat{\psi} - \psi_*)' \left. \frac{\partial^2 g_i}{\partial \psi \partial \psi'} \right|_{\psi_{[i]}} (\hat{\psi} - \psi_*) \\ &= \eta_i + \delta_i \end{aligned}$$

with η_i, δ_i defined in obvious ways, where $\psi_{[i]}$ lies between ψ_* and $\hat{\psi}$. By (S2.9), we have

$$|\eta_i| \le (c/\lambda)(|y_i|\vee|x_i|)(|\xi|/m), \quad |\delta_i| \le (c/\lambda^2)\{\log(m+1)\}^K(|y_i|\vee|x_i|)(|\xi|/m)^2,$$

using the regularization of $\hat{\psi}$ for the latter inequality. Now write

$$\mathrm{E}\{(\hat{\theta}_i-\tilde{\theta}_{i*})^2 1_{\mathcal{A}_\delta\cap\mathcal{A}_\delta\cap\mathcal{C}_\delta}\} = \mathrm{E}(\eta_i^2 1_{\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta})+2\mathrm{E}(\eta_i\delta_i 1_{\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta})+\mathrm{E}(\delta_i^2 1_{\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta}). \tag{S2.10}$$

By the inequalities above (S2.10), the Cauchy-Schwarz inequality, and the Marcinkiewicz-Zygmund inequality, it can be shown that the second and third terms on the right side of (S2.10) are both $o(m^{-1})$. Thus, we have

$$\mathrm{E}\{(\hat{\theta}_i - \tilde{\theta}_{i*})^2 1_{\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta}\} = \mathrm{E}(\eta_i^2 1_{\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta}) + o(m^{-1}). \tag{S2.11}$$

Furthermore, write $h_i = -\partial g_i/\partial\psi'|_{\psi_*}$, and $\mathcal{E} = \mathcal{A}_\delta \cap \mathcal{B}_\delta \cap C_\delta$. It follows from (S2.8) that

$$\eta_i = h_i G^{-1}\xi + h_i G^{-1}\eta(\hat{\psi} - \psi_*) + (h_i/2)G^{-1}[\cdots](\hat{\psi} - \psi_*) = \zeta_i + \delta_{1i} + \delta_{2i}$$

with $\zeta_i, \delta_{1i}, \delta_{2i}$, and $[\cdots]$ defined in obvious ways. Thus, we have

$$\begin{aligned}\mathrm{E}(\eta_i^2 1_{\mathcal{E}}) &= \mathrm{E}(\zeta_i^2 1_{\mathcal{E}}\} + 2\mathrm{E}\{\zeta_i(\delta_{1i} + \delta_{2i})1_{\mathcal{E}}\} + \mathrm{E}\{(\delta_{1i} + \delta_{2i})^2 1_{\mathcal{E}}\} \\ &= \mathrm{E}(\zeta_i^2) - \mathrm{E}(\zeta_i^2 1_{\mathcal{E}^c}) + 2\mathrm{E}\{\zeta_i(\delta_{1i} + \delta_{2i})1_{\mathcal{E}}\} + \mathrm{E}\{(\delta_{1i} + \delta_{2i})^2 1_{\mathcal{E}}\}. \quad \text{(S2.12)}\end{aligned}$$

Note that, on $\mathcal{E}$, we have $|\delta_{i1}| \le |h_i|\cdot\|G^{-1}\|\cdot\|\eta\|\cdot|\hat{\psi}-\psi_*| \le (c/\lambda m^2)|h_i|\cdot\|\eta\|\cdot|\xi|$, using (S2.9). It follows that

$$\mathrm{E}(\delta_{i1}^2 1_{\mathcal{E}}) \le \frac{c}{\lambda^2 m^4}\mathrm{E}(h_i^2\|\eta\|^2|\xi|^2) \le \frac{c}{\lambda^2 m^4}\sum_{r,s,t}\mathrm{E}\left(h_i^2\left|\sum_{j=1}^m u_{jrs}\right|^2\left|\sum_{j=1}^m u_{jt}\right|^2\right), \tag{S2.13}$$

where $u_{jt} = \partial Q_j/\partial\psi_t|_{\psi_*}$ and $u_{jrs} = \partial^2 Q_j/\partial\psi_r\partial\psi_s|_{\psi_*} - \mathrm{E}(\partial^2 Q_j/\partial\psi_r\partial\psi_s|_{\psi_*})$.

We have

$$\begin{aligned}
& \mathrm{E}\left(h_i^2\left|\sum_{j=1}^m u_{jrs}\right|^2\left|\sum_{j=1}^m u_{jt}\right|^2\right) \\
\le\ & c\mathrm{E}\left\{h_i^2\left(u_{irs}^2 + \left|\sum_{j\ne i} u_{jrs}\right|^2\right)\left(u_{it}^2 + \left|\sum_{j\ne i} u_{jt}\right|^2\right)\right\} \\
=\ & c\left[\mathrm{E}(h_i^2 u_{irs}^2 u_{it}^2) + \mathrm{E}(h_i^2 u_{irs}^2)\mathrm{E}\left(\sum_{j\ne i} u_{jt}\right)^2\right. \\
& \left. + \mathrm{E}(h_i^2 u_{it}^2)\mathrm{E}\left(\sum_{j\ne i} u_{jrs}\right)^2 + \mathrm{E}(h_i^2)\mathrm{E}\left\{\left(\sum_{j\ne i} u_{jrs}\right)^2\left(\sum_{j\ne i} u_{jt}\right)^2\right\}\right] \\
\le\ & c\left[m + \mathrm{E}\left\{\left(\sum_{j\ne i} u_{jrs}\right)^2\left(\sum_{j\ne i} u_{jt}\right)^2\right\}\right].
\end{aligned} \tag{S2.14}$$

By (S2.13), (S2.14), and the Cauchy-Schwarz and Marcinkiewicz-Zygmund inequalities, it can be shown that $\mathrm{E}(\delta_{i1}^2 1_{\mathcal{E}}) = O(m^{-2})$. Also, we have, on $\mathcal{E}$,

$$|\delta_{i2}| \le c|h_i|\cdot|\hat\psi - \psi_*|^2 \le \frac{c}{\lambda^2 m^2}|h_i|\cdot|\xi|^2 = \frac{c}{\lambda^2 m^2}|h_i|\sum_t\left|\sum_{j=1}^m u_{jt}\right|^2.$$

Thus, similarly, it can be shown that $\mathrm{E}(\delta_{i2}^2) = O(m^{-2})$. Furthermore, we have $|\zeta_i| \le (c/m)|h_i|\cdot|\xi|$. Thus, similarly, it can be shown that $\mathrm{E}(\zeta_i^w) = O(m^{-w/2})$, $w = 2, 4$.

It follows that the fourth term on the right side of (S2.12) is $O(m^{-2})$; the third term is $O(m^{-3/2})$ (by the Cauchy-Schwarz inequality); and the second term is $O\{m^{-(1+d/2)}\}$ [again by the Cauchy-Schwarz inequality, and a bound on $\mathrm{P}(\mathcal{E}^c)$ obtained earlier]. (S2.5) now follows by combining (S2.6), (S2.7), (S2.11) and (S2.12).

Recall the definition of u_j below (3.6) of LMJ20. By (2.10) of LMJ20, it can be shown that $\mathrm{E}(u_j) = 0$ for every $1 \leq j \leq m$. Therefore, by independence, we have

$$
\begin{aligned}
& \mathrm{E}\left(\frac{\partial g_i^*}{\partial \psi'} G^{-1}\xi\right)^2 \\
= \quad & \mathrm{E}\left\{h_i G^{-1}\left(u_i + \sum_{j\neq i} u_j\right)\right\}^2 \\
= \quad & \mathrm{E}(h_i G^{-1} u_i)^2 + 2\mathrm{E}\{(h_i G^{-1} u_i) h_i\} G^{-1} \mathrm{E}\left(\sum_{j\neq i} u_j\right) + \mathrm{E}\left(h_i G^{-1} \sum_{j\neq i} u_j\right)^2 \\
= \quad & O(m^{-2}) + \sum_{j\neq i} \mathrm{E}(h_i G^{-1} u_j)^2 \\
= \quad & \sum_{j=1}^{m} \mathrm{E}(u_j' G^{-1} R_i G^{-1} u_j) + o(m^{-1}) \\
= \quad & \mathrm{E}(d_i) + o(m^{-1}), \qquad \text{(S2.15)}
\end{aligned}
$$

where $R_i = \mathrm{E}(h_i' h_i) = r_i^2 \mathrm{diag}\{\mathrm{E}(x_1 x_1'), (A_* + D_i)^{-1}\}$ with $r_i = D_i/(A_* + D_i)$, and

$$
d_i = r_i^2 \left\{\frac{u_{0m}}{s_{0m}^2} + \frac{V_{0m}}{(A_* + D_i) s_{1m}^2}\right\}.
$$

Hereafter, $s_{km}, k = 0, 1, 2$, etc. are $\hat{s}_{km}, k = 0, 1, 2$, etc.defined above (3.2) of LMJ20 with $\hat{\psi}$ replaced by ψ_*. The last equation in (S2.15) can be derived after some further computation. Let $\hat{d}_i$ be d_i with ψ_* replaced by $\hat{\psi}$. We show that

$$\mathrm{E}(\hat{d}_i - d_i) = o(m^{-1}). \tag{S2.16}$$

To show (S2.16), write, similar to (S2.6),

$$\mathrm{E}(\hat{d}_i - d_i) = \mathrm{E}\{(\hat{d}_i - d_i)1_{\mathcal{A}_\delta \cap \mathcal{B}_\delta \cap \mathcal{C}_\delta \cap \mathcal{D}}\} + \mathrm{E}\{(\hat{d}_i - d_i)1_{\mathcal{A}_\delta^c \cup \mathcal{B}_\delta \cup \mathcal{C}_\delta^c \cup \mathcal{D}^c}\}. \tag{S2.17}$$

By a similar argument as that leading to (S2.7) (using Hölder's inequality instead of Cauchy-Schwarz inequality), it can be shown that the second term on the right side of (S2.17) is $o(m^{-1})$. Furthermore, it can be shown, using similar arguments as those leading to (S2.15) [involving Taylor expansion, (S2.9), and Hölder's inequality instead Cauchy-Schwarz inequality], it can be shown that the first term on the right side of (S2.17) is also $o(m^{-1})$. (S2.16) thus follows. Combining (S2.5), (S2.15), (S2.16), we have

$$I_2 = \mathrm{E}(\hat{\theta}_i - \tilde{\theta}_{i*})^2 = \mathrm{E}(\hat{d}_i) + o(m^{-1}). \tag{S2.18}$$

Now let us consider I_1. Similar to (S2.6), (S2.7), we have

$$\begin{aligned} I_1 &= \mathrm{E}\{(\tilde{\theta}_{i*} - \theta_i)(\hat{\theta}_i - \tilde{\theta}_{i*})1_{\mathcal{A}_\delta \cap \mathcal{B}_\delta \cap \mathcal{C}_\delta}\} + \mathrm{E}\{(\tilde{\theta}_{i*} - \theta_i)(\hat{\theta}_i - \tilde{\theta}_{i*})1_{\mathcal{A}_\delta^c \cup \mathcal{B}_\delta^c \cup \mathcal{C}_\delta^c}\} \\ &= I_{11} + I_{12}, \end{aligned} \tag{S2.19}$$

with I_{11}, I_{12} defined in obvious ways, and

$$
\begin{aligned}
|I_{12}| &\leq \left[\mathrm{E}\{(\tilde{\theta}_{i*}-\theta_i)^2(\hat{\theta}_i-\tilde{\theta}_{i*})^2\}\right]^{1/2}\mathrm{P}(\mathcal{A}_\delta^c\cup\mathcal{B}_\delta^c\cup\mathcal{C}_\delta^c)^{1/2} \\
&\leq \left[\mathrm{E}\{(\tilde{\theta}_{i*}-\theta_i)^4\}\right]^{1/4}\left[\mathrm{E}\{(\hat{\theta}_i-\tilde{\theta}_{i*})^4\}\right]^{1/4}\mathrm{P}(\mathcal{A}_\delta^c\cup\mathcal{B}_\delta^c\cup\mathcal{C}_\delta^c)^{1/2} \\
&= o(m^{-1}). \qquad \text{(S2.20)}
\end{aligned}
$$

Next (recall $\mathcal{E}=\mathcal{A}_\delta\cap\mathcal{B}_\delta\cap\mathcal{C}_\delta$), we have, by Taylor series expansion,

$$
\begin{aligned}
\hat{\theta}_i-\tilde{\theta}_{i*} &= g_i(\hat{\psi},y_i)-g_i(\psi_*,y_i) \\
&= \left.\frac{\partial g_i}{\partial\psi'}\right|_{\psi_*}(\hat{\psi}-\psi_*)+\frac{1}{2}(\hat{\psi}-\psi_*)'\left.\frac{\partial^2 g_i}{\partial\psi\partial\psi'}\right|_{\psi_*}(\hat{\psi}-\psi_*)+\rho_i, \quad \text{(S2.21)}
\end{aligned}
$$

where ρ_i denotes the remaining term in the second-order Taylor expansion. It can be shown, by (S2.9) (also see above *A3* of LMJ20), that, on $\mathcal{E}$, we have

$$
|(\tilde{\theta}_{i*}-\theta_i)\rho_i|\leq c\{\log(m+1)\}^\gamma(|y_i|+|x_i|+|\theta_i|)^2\left(\frac{|\xi|}{m}\right)^3.
$$

Therefore, by Hölder' and Marcinkiewicz-Zygmund inequalities, we have

$$
\begin{aligned}
\mathrm{E}\left\{|(\tilde{\theta}_{i*}-\theta_i)\rho_i|1_{\mathcal{E}}\right\} &\leq c\{\log(m+1)\}^\gamma\left[\mathrm{E}\{(|y_i|+|x_i|+|\theta_i|)^8\}\right]^{1/4}\left\{\mathrm{E}\left(\frac{|\xi|}{m}\right)^4\right\}^{3/4} \\
&\leq c\{\log(m+1)\}^\gamma m^{-3/2} \\
&= o(m^{-1}).
\end{aligned}
$$

Therefore, combining with (S2.21), we have

$$\begin{aligned} I_{11} &= \mathrm{E}\left\{(\tilde{\theta}_{i*}-\theta_i)\left.\frac{\partial g_i}{\partial \psi'}\right|_{\psi_*}(\hat{\psi}-\psi_*)1_{\mathcal{E}}\right\} \\ &\quad +\frac{1}{2}\mathrm{E}\left\{(\tilde{\theta}_{i*}-\theta_i)(\hat{\psi}-\psi_*)'\left.\frac{\partial^2 g_i}{\partial\psi\partial\psi'}\right|_{\psi_*}(\hat{\psi}-\psi_*)1_{\mathcal{E}}\right\}+o(m^{-1}) \\ &= w_1+\frac{w_2}{2}+o(m^{-1}) \qquad \text{(S2.22)} \end{aligned}$$

with w_1, w_2 defined in obvious ways. Let us first consider w_2. By (S2.8), we have

$$\hat{\psi}-\psi_* = -G^{-1}\xi - G^{-1}\left(\eta+\frac{1}{2}[\cdots]\right)(\hat{\psi}-\psi_*). \qquad \text{(S2.23)}$$

By the definition of $\mathcal{B}_\delta$, and (S2.9), it can be shown that

$$\left|G^{-1}\left(\eta+\frac{1}{2}[\cdots]\right)(\hat{\psi}-\psi_*)\right| \le \frac{c}{m^2}(\|\eta\|+|\xi|)|\xi|$$

on $\mathcal{E}$. It can then be shown that

$$\begin{aligned} w_2 &= \mathrm{E}\left\{(\tilde{\theta}_{i*}-\theta_i)\xi' G^{-1}\left.\frac{\partial^2 g_i}{\partial\psi\partial\psi'}\right|_{\psi_*}G^{-1}\xi 1_{\mathcal{E}}\right\}+o(m^{-1}) \\ &= \mathrm{E}\left\{(\tilde{\theta}_{i*}-\theta_i)\xi' G^{-1}\left.\frac{\partial^2 g_i}{\partial\psi\partial\psi'}\right|_{\psi_*}G^{-1}\xi\right\}+o(m^{-1}). \qquad \text{(S2.24)} \end{aligned}$$

Write $P_i = \partial^2 g_i/\partial\psi\partial\psi'|_{\psi_*}(\tilde{\theta}_{i*}-\theta_i)$. It is easy to see that $\mathrm{E}(u_j' G^{-1}P_i G^{-1}u_k) = 0$, if $j \neq k$. Furthermore, by (2.10) of LMJ20, it can be shown that, for $j \neq i$, we have $\mathrm{E}(u_j' G^{-1}P_i G^{-1}u_j) = \mathrm{E}\{\mathrm{tr}(G^{-1}P_i G^{-1}u_j u_j')\} = \mathrm{tr}\{G^{-1}\mathrm{E}(P_i)G^{-1}\mathrm{E}(u_j u_j')\}$

$= 0$, hence

$$\begin{aligned} \mathrm{E}\left\{(\tilde{\theta}_{i*}-\theta_i)\xi' G^{-1}\left.\frac{\partial^2 g_i}{\partial\psi\partial\psi'}\right|_{\psi_*} G^{-1}\xi\right\} &= \sum_{j,k=1}^m \mathrm{E}(u_j' G^{-1} P_i G^{-1} u_k) \\ &= \sum_{j=1}^m \mathrm{E}(u_j' G^{-1} P_i G^{-1} u_j) \\ &= \mathrm{E}(u_i' G^{-1} P_i G^{-1} u_i) \\ &= O(m^{-2}). \end{aligned} \quad \text{(S2.25)}$$

Combining (S2.24), (S2.25), we have $w_2 = o(m^{-1})$.

Now consider w_1. Write $M_s = \partial^3 Q/\partial\psi_s\partial\psi\partial\psi'|_{\psi_{[s]}}$, $M_{s*} = M_s$ with $\psi_{[s]}$ replaced by ψ_*, and $H_s = \mathrm{E}(M_{s*})$. Going back to (S2.23), it is seen, by Taylor series expansion, that

$$\begin{aligned} & [\cdots](\hat{\psi}-\psi_*) \\ = \ & [(\hat{\psi}-\psi_*)' M_s(\hat{\psi}-\psi_*)]_{1\le s\le p+1} = [(\hat{\psi}-\psi_*) M_{s*}(\hat{\psi}-\psi_*)]_{1\le s\le p+1} + \rho_1 \\ = \ & [(\hat{\psi}-\psi_*) H_s(\hat{\psi}-\psi_*)]_{1\le s\le p+1} + [(\hat{\psi}-\psi_*)(M_{s*}-H_s)(\hat{\psi}-\psi_*)]_{1\le s\le p+1} + \rho_1 \\ = \ & [(\hat{\psi}-\psi_*) H_s(\hat{\psi}-\psi_*)]_{1\le s\le p+1} + \rho_2 + \rho_1, \end{aligned}$$

where ρ_1, ρ_2 satisfy that, on $\mathcal{E}$ [see (S2.9)],

$$|\rho_1| \le c\{\log(m+1)\}^{\gamma}\left\{\sum_{j=1}^m (y_j^2 + |x_j|^2 + 1)\right\}\left(\frac{|\xi|}{m}\right)^3,$$

$$|\rho_2| \le \|M_{s*} - H_s\|\left(\frac{|\xi|}{m}\right)^2.$$

Next, by (S2.23), it can be shown that, on $\mathcal{E}$,

$$(\hat{\psi}-\psi_*)H_s(\hat{\psi}-\psi_*) = (-\xi' G^{-1} - \cdots')H_s(-G^{-1}\xi - \cdots) = \xi' G^{-1} H_s G^{-1}\xi + \rho_{3s},$$

where ρ_{3s} $(1 \leq s \leq p+1)$ satisfies

$$|\rho_{3s}| \leq c\left\{\frac{1}{m}\sum_{j=1}^{m}(y_j^2 + |x_j|^2 + 1)\right\}\left(1 + \frac{|\xi| + \|\eta\|}{m}\right)(|\xi| + \|\eta\|)\left(\frac{|\xi|}{m}\right)^2.$$

Similarly, on $\mathcal{E}$ we have $G^{-1}\eta(\hat{\psi}-\psi_*) = G^{-1}\eta(-G^{-1}\xi - \cdots) = -G^{-1}\eta G^{-1}\xi + \rho_4$ with

$$|\rho_4| \leq \frac{c}{m^3}(|\xi| + \|\eta\|)|\xi| \cdot \|\eta\|.$$

Combining (S2.23), and the above results, we obtained a more detailed expansion:

$$\begin{aligned}\hat{\psi} - \psi_* &= -G^{-1}\xi + G^{-1}\eta G^{-1}\xi - \frac{1}{2}G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\leq s\leq p+1} \\ &\quad -\frac{1}{2}G^{-1}(\rho_1 + \rho_2 + \rho_3) - \rho_4, \end{aligned} \tag{S2.26}$$

where $\rho_3 = (\rho_{3s})_{1\leq s\leq p+1}$. Write $q_i = (\tilde{\theta}_{i*} - \theta_i)(\partial g_i/\partial\psi'|_{\psi_*})$. It can then be shown that

$$\begin{aligned} w_1 &= -\mathrm{E}\left[q_i\left\{G^{-1}\xi - G^{-1}\eta G^{-1}\xi + \frac{1}{2}G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\leq s\leq p+1}\right\}1_{\mathcal{E}}\right] + o(m^{-1}) \\ &= -\mathrm{E}\left[q_i\left\{G^{-1}\xi - G^{-1}\eta G^{-1}\xi + \frac{1}{2}G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\leq s\leq p+1}\right\}\right] + o(m^{-1}). \end{aligned}$$

Furthermore, we have $\mathrm{E}(q_i G^{-1}\xi) = \sum_{j=1}^{m}\mathrm{E}(q_i G^{-1}u_j) = \mathrm{E}(q_i G^{-1}u_i)$. Next, write $V_j = \partial^2 Q_j/\partial\psi\partial\psi'|_{\psi_*}$. By similar arguments as in (S2.25), we have

$$\begin{aligned} \mathrm{E}\left(q_i G^{-1} \left.\frac{\partial^2 Q}{\partial\psi\partial\psi'}\right|_{\psi_*} G^{-1}\xi\right) &= \sum_{j,k=1}^{m} \mathrm{E}(q_i G^{-1} V_j G^{-1} u_k) \\ &= \mathrm{E}(q_i G^{-1} V_i G^{-1} u_i) = O(m^{-2}). \end{aligned}$$

Finally, let F_s denote the sth column of G^{-1}. Then, we have

$$q_i G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\le s\le p+1} = \sum_{s=1}^{p+1} q_i F_s \xi' G^{-1} H_s G^{-1}\xi$$

$$= \sum_{s=1}^{p+1}\sum_{j,k=1}^{m} q_i F_s u_j' G^{-1} H_s G^{-1} u_k, \quad \text{hence}$$

$$\mathrm{E}\left\{q_i G^{-1}(\xi' G^{-1} H_s G^{-1}\xi)_{1\le s\le p+1}\right\} = \sum_{s=1}^{p+1}\sum_{j,k=1}^{m} \mathrm{E}(q_i F_s u_j' G^{-1} H_s G^{-1} u_k)$$

$$= \sum_{j=1}^{p+1} \mathrm{E}(q_i F_s u_i' G^{-1} H_s G^{-1} u_i) = \mathrm{E}\left\{q_i G^{-1}(u_i' G^{-1} H_s G^{-1} u_i)_{1\le i\le p+1}\right\} = O(m^{-2}).$$

Therefore, combining the above results, we have $w_1 = -\mathrm{E}(q_i G^{-1} u_i) + o(m^{-1})$. Combining the results of (S2.19), (S2.20), (S2.22), and those regarding w_1 and w_2, we conclude that

$$\begin{aligned} I_1 &= \mathrm{E}\{(\tilde\theta_{i*} - \theta_i)(\hat\theta_i - \tilde\theta_{i*})\} \\ &= -\mathrm{E}(q_i G^{-1} u_i) + o(m^{-1}) \\ &= -\mathrm{E}\left[\frac{r_i^4}{s_{1m}}\left\{\frac{(y_i - x_i'\beta_*)^4}{(A_* + D_i)^2} - 3\right\}\right] + o(m^{-1}) \\ &\quad = -\mathrm{E}(g_i) + o(m^{-1}), \end{aligned} \tag{S2.27}$$

with g_i defined in an obvious way (recall s_{1m} is $\hat s_{1m}$ with $\hat\psi$ replaced by ψ_*). The second to last expression on the right side of (S2.27) can be derived with some algebra.

It can be seen that $\mathrm{E}(g_i)$ is zero when there is no model misspecification; otherwise, this term may not vanish. In fact, by (3.4) of LMJ20, it is seen that $\mathrm{E}(g_i)$ is associated with the kurtosis of δ_i, defined below (2.1) of LMJ20. Directly applying the observed information idea to $\mathrm{E}(g_i)$ would suggest removal of the E and replacement of ψ_* by $\hat{\psi}$ in g_i. This would produce a term involving a single observation, y_i, which has large variation; in fact, this was an approach used by Jiang *et al.* (2011), which we intend to avoid. Instead, we apply identity (3.5) of LMJ20 to replace $\mathrm{E}(g_i)$ by a term that involves all of the observations, that is, $\mathrm{E}(a_i)$, where a_i is $\hat{a}_i$ defined above (3.2) of LMJ20 with $\hat{\psi}$ replaced by ψ_*.

Furthermore, by similar arguments as those leading to (S2.16), it can be shown that $\mathrm{E}(\hat{a}_i - a_i) = o(m^{-1})$. It follows that

$$I_1 = -\mathrm{E}(\hat{a}_i) + o(m^{-1}). \tag{S2.28}$$

Finally, let us deal with the leading term on the right side of (S2.4). Write

$$c_i(A_*) = \frac{A_* D_i}{A_* + D_i} = D_i - \frac{D_i^2}{A_* + D_i}. \tag{S2.29}$$

Recall $\mathcal{E} = \mathcal{A}_\delta \cap \mathcal{B}_\delta \cap \mathcal{C}_\delta$ and, by earlier results, $\mathrm{P}(\mathcal{E}^c) = O(m^{-d})$. Thus, we have

$$\mathrm{E}\left(\frac{1}{\hat{A} + D_i}\right) = \mathrm{E}\left(\frac{1_{\mathcal{E}}}{\hat{A} + D_i}\right) + o(m^{-1}). \tag{S2.30}$$

Using an elementary expansion (see Jiang 2010, p. 103), we have

$$\frac{1}{\hat{A}+D_i}=\frac{1}{A_*+D_i}-\frac{\hat{A}-A_*}{(A_*+D_i)^2}+\frac{(\hat{A}-A_*)^2}{(A_*+D_i)^3}-\frac{(\hat{A}-A_*)^3}{(A_*+D_i)^3(\hat{A}+D_i)}.$$

Applying (S2.9), it can be shown that

$$\mathrm{E}\left\{\frac{(\hat{A}-A_*)^3 1_{\mathcal{E}}}{(A_*+D_i)^3(\hat{A}+D_i)}\right\}\le c\mathrm{E}\left(\frac{|\xi|}{m}\right)^3=o(m^{-1}).$$

Therefore, we have

$$\begin{aligned}\mathrm{E}\left(\frac{1_{\mathcal{E}}}{\hat{A}+D_i}\right) &= \frac{\mathrm{P}(\mathcal{E})}{A_*+D_i}-\frac{\mathrm{E}\{(\hat{A}-A_*)1_{\mathcal{E}}\}}{(A_*+D_i)^2}+\frac{\mathrm{E}\{(\hat{A}-A_*)^2 1_{\mathcal{E}}\}}{(A_*+D_i)^3}+o(m^{-1})\\ &= \frac{1}{A_*+D_i}-\frac{\mathrm{E}\{(\hat{A}-A_*)1_{\mathcal{E}}\}}{(A_*+D_i)^2}+\frac{\mathrm{E}\{(\hat{A}-A_*)^2 1_{\mathcal{E}}\}}{(A_*+D_i)^3}+o(m^{-1}).\end{aligned}$$

Furthermore, by (S2.26), we have

$$\begin{aligned}\hat{A}-A_* &= -h'\xi+h'\eta G^{-1}\xi-\frac{1}{2}h'(\xi' G^{-1}H_s G^{-1}\xi)_{1\le s\le p+1}\\ &\quad -\frac{1}{2}h'(\rho_1+\rho_2+\rho_3)-(0'\ 1)\rho_4,\end{aligned}$$

where $h'=(0'\ 1)G^{-1}=(0'\ h_{22})$ with $G^{-1}=\mathrm{diag}(H_{11},h_{22})$. By earlier arguments, we have $\mathrm{E}\{h'(\rho_1+\rho_2+\rho_3)1_{\mathcal{E}}\}=o(m^{-1})$, $\mathrm{E}\{(0'\ 1)\rho_4 1_{\mathcal{E}}\}=o(m^{-1})$. Also, we have $\mathrm{E}(\xi)=0$, $h'\mathrm{E}(\eta G^{-1}\xi 1_{\mathcal{E}^c})=o(m^{-1})$, $h'\mathrm{E}\{(\xi' G^{-1}H_s G^{-1}\xi)_{1\le s\le p+1}1_{\mathcal{E}^c}\}=o(m^{-1})$. It follows that $\mathrm{E}\{(\hat{A}-A_*)1_{\mathcal{E}}\}=h'\mathrm{E}(\eta G^{-1}\xi)-(1/2)h'[\mathrm{E}(\xi' G^{-1}H_s G^{-1}\xi)]_{1\le s\le p+1}$ $+o(m^{-1})$. Next, by (A.23) and the inequality below it, it can be shown that

$$\mathrm{E}((\hat{A}-A_*)^2 1_{\mathcal{E}}\}=\mathrm{E}(\xi' hh'\xi)+o(m^{-1}).$$

Combining the above results, we conclude that

$$
\begin{aligned}
\mathrm{E}\{c_i(\hat{A})\} &= D_i\left\{1 - D_i\mathrm{E}\left(\frac{1}{\hat{A}+D_i}\right)\right\} \\
&= D_i\left\{1 - \frac{D_i}{A_*+D_i} + \frac{D_i}{(A_*+D_i)^2}h'\mathrm{E}(\eta G^{-1}\xi)\right. \\
&\quad -\frac{D_i}{2(A_*+D_i)^2}h'\left[\mathrm{E}(\xi' G^{-1}H_s G^{-1}\xi)\right]_{1\le s\le p+1} \\
&\quad \left. -\frac{D_i}{(A_*+D_i)^3}\mathrm{E}(\xi' h h'\xi) + o(m^{-1})\right\} \\
&= c_i(A_*) + r_i^2\left\{h'\mathrm{E}(\eta G^{-1}\xi) - \frac{1}{2}h'\left[\mathrm{E}(\xi' G^{-1}H_s G^{-1}\xi)\right]_{1\le s\le p+1}\right. \\
&\quad \left. -\frac{\mathrm{E}(\xi' h h'\xi)}{A_*+D_i}\right\} + o(m^{-1}). \qquad \text{(S2.31)}
\end{aligned}
$$

Furthermore, it is easy to see, by (2.10) of LMJ20, that $\mathrm{E}(\eta G^{-1}\xi) = \sum_{j=1}^m \mathrm{E}(V_j G^{-1}u_j)$, where V_i is defined below (S2.26). Further derivations show that

$$
\begin{aligned}
h'\mathrm{E}(V_j G^{-1}u_j) &= -4s_{1m}^{-1}\mathrm{E}\left\{\frac{D_j^4}{(A_*+D_j)^5}x_j' H_{11}x_j(y_j - x_j'\beta_*)^2\right\} \\
&\quad -3s_{1m}^{-2}\mathrm{E}\left\{\frac{D_j^4}{(A_*+D_j)^5}\left\{\frac{(y_j - x_j'\beta_*)^4}{(A_*+D_j)^2} - 1\right\}\right]
\end{aligned}
$$

[note $G^{-1} = \mathrm{diag}(H_{11}, h_{22})$]. It then follows that we have the expression

$$
h'\mathrm{E}(\eta G^{-1}\xi) = -\mathrm{E}\left(\frac{2u_{1m}}{s_{0m}s_{1m}} + \frac{3V_{1m}}{s_{1m}^2}\right), \qquad \text{(S2.32)}
$$

using identity (3.3) of LMJ20. Similarly, we have

$$
[\mathrm{E}(\xi' G^{-1}H_s G^{-1}\xi)]_{1\le s\le p+1} = \sum_{j=1}^m \mathrm{E}(u_j' G^{-1}H_s G^{-1}u_j)_{1\le s\le p+1}.
$$

Thus, we have $h'[\mathrm{E}(\xi' G^{-1}H_s G^{-1}\xi)]_{1\le s\le p+1} = \sum_{j=1}^m h_{22}\mathrm{E}(u_j' G^{-1}H_{p+1}G^{-1}u_j)$.

Also, the following expression can be derived:

$$H_{p+1} = \mathrm{E}\left(\frac{\partial^3 Q}{\partial A \partial \psi \partial \psi'}\bigg|_{\psi_*} \right) = -4 \begin{bmatrix} s_{1m}\mathrm{E}(x_1 x_1') & 0 \\ 0 & 3s_{2m} \end{bmatrix}.$$

Therefore, the following expression can then be derived:

$$h'[\mathrm{E}(\xi' G^{-1} H_s G^{-1} \xi)]_{1 \le s \le p+1} = -2\mathrm{E}\left\{ \frac{u_{0m}}{s_{0m}^2} + 3\left(\frac{s_{2m} V_{0m}}{s_{1m}^3} \right) \right\}. \qquad \text{(S2.33)}$$

Similarly, it can be shown that

$$\mathrm{E}(\xi' h h' \xi) = \mathrm{E}\left(\frac{V_{0m}}{s_{1m}^2} \right). \qquad \text{(S2.34)}$$

Combining (S2.31)–(S2.34), we have

$$\begin{aligned} & \mathrm{E}\{c_i(\hat{A})\} \\ = \;& c_i(A_*) - \mathrm{E}\left[r_i^2 \left\{ \frac{2u_{1m}}{s_{0m}s_{1m}} + \frac{3}{s_{1m}^3}(s_{1m}V_{1m} - s_{2m}V_{0m}) - \frac{u_{0m}}{s_{0m}^2} + \frac{V_{0m}}{(A_* + D_i)s_{1m}^2} \right\} \right] \\ & + o(m^{-1}) \\ = \;& c_i(A_*) - \mathrm{E}(h_i) + o(m^{-1}), \qquad \text{(S2.35)} \end{aligned}$$

with h_i defined in an obvious way, where u_{km}, V_{km} are $\hat{u}_{km}, \hat{V}_{km}$ defined above (3.2) of LMJ20, respectively, with $\hat{\psi}$ replaced by ψ_*. Then, as argued before, it can be shown that

$$\mathrm{E}(\hat{h}_i - h_i) = o(m^{-1}), \qquad \text{(S2.36)}$$

where $\hat{h}_i$ is h_i with ψ_* replaced by $\hat{\psi}$.

Combining (S2.4), (S2.18), (S2.28), (S2.35), and (S2.36), we conclude that

$$\mathrm{MSPE}(\hat{\theta}_i) = \mathrm{E}\{c_i(\hat{A}) + \hat{h}_i - 2\hat{a}_i + \hat{d}_i\} + o(m^{-1}) = \mathrm{E}\{\widehat{\mathrm{MSPE}}(\hat{\theta}_i)\} + o(m^{-1}),$$

where $\hat{d}_i + \hat{h}_i = \hat{b}_i$, which has the expression given above (3.2) of LMJ20.

S3 Computational details, additional tables and figure

First, we present computational details of the simulation study in algorithmic form.

Input: $m \in \{20, 40, 80, 160, 320, 640\}$, $K = 10000$, and Model ID $\in \{(I), (II), (III)\}$. Step 1. Set $D_i = (i-1)/(m-1) + 0.5$, $i = 1, 2, \cdots, m-1$, then fix them throughout the simulations.

Step 2. For each $k = 1, 2, \cdots, K$, run:

(i) Generate the random samples $\{x_i^{(k)}, y_i^{(k)}\}_{i=1}^m$ from Model ID. (We have put the R codes of these three different models in files entitled Example_1.R, Example_2.R and Example_3.R in the supplement, respectively.)

(ii) Store $\{\theta_i^{(k)}\}_{i=1}^m$, where $\theta_i^{(k)} = y_i - e_i$. Note that $\{\theta_i^{(k)}\}_{i=1}^m$ are used only in approximating the true MSPE.

(iii) Compute the MPR, PR, Naive, and JNR estimates based on the data set $\{x_i^{(k)}, y_i^{(k)}, D_i\}_{i=1}^m$. Store them in $\{\hat{\theta}_{i,\mathrm{MPR}}^{(k)}\}_{i=1}^m$, $\{\hat{\theta}_{i,\mathrm{PR}}^{(k)}\}_{i=1}^m$, $\{\hat{\theta}_{i,\mathrm{Nai}}^{(k)}\}_{i=1}^m$ and $\{\hat{\theta}_{i,\mathrm{JNR}}^{(k)}\}_{i=1}^m$, respectively.

(iv) Store $\text{PoNE}_{\text{JNR}}^{(k)} = \#\{i : \hat{\theta}_{i,\text{JNR}}^{(k)} < 0\}/m$ and $\text{PoNE}_{\text{Nai}}^{(k)} = \#\{i : \hat{\theta}_{i,\text{Nai}}^{(k)} < 0\}/m$, where $\#(A)$ denotes the cardinal number of a set A.

Step 3. Average the MPR, PR, Naive, and JNR estimates over these $K = 10000$ computations. That is, compute

$$\begin{aligned} \overline{\text{PR}}_i &= \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_{i,\text{PR}}^{(k)}, \quad \overline{\text{MPR}}_i = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_{i,\text{MPR}}^{(k)}, \\ \overline{\text{Nai}}_i &= \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_{i,\text{Nai}}^{(k)}, \quad \overline{\text{JNR}}_i = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_{i,\text{JNR}}^{(k)}. \end{aligned}$$

Also compute the approximate true MSPE values by

$$\overline{\text{True}}_i = \frac{1}{K}\sum_{k=1}^{K}(\theta_i^{(k)})^2, \quad i = 1, 2, \cdots, m.$$

Step 4. Compute the RBs of each estimator, i.g., PR, by

$$\text{RB}_i = \frac{\overline{\text{PR}}_i - \overline{\text{True}}_i}{\overline{\text{True}}_i}, \quad i = 1, 2, \cdots, m,$$

and store the corresponding mean and Sd of $\{\text{RB}_i\}_{i=1}^m$.

Output: The mean and Sd of $\{100 \times \text{RB}_i\}_{i=1}^m$, as well as $\{100 \times |\text{RB}_i|\}_{i=1}^m$, of four estimators, and the means of $\{\text{PoNE}_{\text{JNR}}^{(k)}\}_k^K$ and $\{\text{PoNE}_{\text{Nai}}^{(k)}\}_k^K$.

Next, we present additional tables for the simulation study and real-data analysis.

Table A1 presents the standard deviations (s.d.) of the %RB and %|RB| corresponding to Table 1 of LMJ20.

Table A.1: Comparison of MSPE Estimators in Term of s.d. of %RB and %|RB|

Example	n	MPR		PR		Naive		JNR	
		%RB	%|RB|	%RB	%|RB|	%RB	%|RB|	%RB	%|RB|
(I)	20	0.0257	0.0228	0.0255	0.0223	0.1699	0.1699	0.0548	0.0507
	40	0.0152	0.0080	0.0149	0.0073	0.1050	0.1050	0.0198	0.0116
	80	0.0146	0.0095	0.0148	0.0096	0.0647	0.0647	0.0226	0.0135
	160	0.0140	0.0082	0.0140	0.0082	0.0368	0.0366	0.0207	0.0129
	320	0.0138	0.0081	0.0138	0.0081	0.0257	0.0238	0.0209	0.0129
	640	0.0145	0.0089	0.0145	0.0089	0.0231	0.0177	0.0219	0.0138
(II)	20	0.0141	0.0073	0.0139	0.0139	0.1368	0.1368	0.0161	0.0114
	40	0.0139	0.0089	0.0137	0.0078	0.0805	0.0805	0.0195	0.0124
	80	0.0148	0.0093	0.0148	0.0095	0.0444	0.0444	0.0181	0.0111
	160	0.0141	0.0082	0.0142	0.0082	0.0278	0.0263	0.0159	0.0097
	320	0.0149	0.0092	0.0149	0.0092	0.0209	0.0178	0.0185	0.0115
	640	0.0143	0.0085	0.0143	0.0085	0.0181	0.0127	0.0173	0.0105
(III)	20	0.0198	0.0103	0.0205	0.0154	0.1460	0.1460	0.0219	0.0125
	40	0.0152	0.0094	0.0151	0.0089	0.0899	0.0899	0.0177	0.0114
	80	0.0144	0.0080	0.0143	0.0077	0.0516	0.0516	0.0178	0.0106
	160	0.0150	0.0091	0.0150	0.0090	0.0317	0.0313	0.0189	0.0109
	320	0.0139	0.0085	0.0139	0.0085	0.0222	0.0189	0.0187	0.0116
	640	0.0140	0.0084	0.0140	0.0084	0.0191	0.0138	0.0177	0.0111

Table A.2 present the raw data and analysis results for Subsection 5.1 of LMJ20. It is divided into Part I and Part II due to its length.

Table A.2: Raw Data, OBP, and MSPE Estimates for Poverty Ratio; Part I

Index	State	y_i	x_{1i}	x_{2i}	x_{3i}	$\sqrt{D_i}$	OBP	MPR	PR	JNR
1	Alabama	18.332	22.721	12.329	0.711	3.541	19.231	4.155	3.820	−5.791
2	Alaska	7.859	12.161	11.974	−0.893	3.140	9.460	3.907	3.676	−1.684
3	Arizona	12.906	19.634	17.811	−0.477	2.892	15.869	3.718	3.557	6.646
4	Arkansas	15.069	24.438	13.553	0.959	3.596	20.025	4.184	3.836	19.632
5	California	19.368	20.254	15.887	−0.973	1.390	18.408	1.866	1.851	1.314
6	Colorado	11.312	13.292	11.893	−0.804	2.707	10.996	3.558	3.449	−2.799
7	Connecticut	9.716	10.520	9.227	2.010	3.148	9.209	3.913	3.680	−3.326
8	Delaware	17.037	14.772	10.572	−0.399	3.741	12.952	4.256	3.875	8.793
9	District of Columbia	22.947	25.760	24.245	3.043	5.273	23.396	4.703	4.101	5.819
10	Florida	16.639	21.304	13.240	0.082	1.835	17.453	2.556	2.591	0.387
11	Geogia	18.330	19.717	14.361	0.357	3.375	17.153	4.061	3.767	−4.959
12	Hawaii	9.526	15.892	12.729	−2.037	3.708	12.307	4.240	3.866	1.152
13	Idaho	17.116	17.367	10.071	−1.528	3.169	14.710	3.927	3.688	1.456
14	Illinois	13.002	15.163	11.677	0.222	1.781	12.865	2.480	2.513	−0.213
15	Indiana	6.873	14.130	8.923	0.337	2.808	10.492	3.648	3.511	10.695
16	Iowa	5.708	11.983	8.580	0.368	2.720	8.756	3.569	3.457	7.144
17	Kansas	16.804	13.208	9.981	0.147	3.024	12.258	3.822	3.624	17.388
18	Kentucky	14.534	20.863	12.917	0.359	3.284	17.129	4.004	3.734	1.344
19	Louisiana	25.852	25.158	15.136	0.112	3.901	22.183	4.327	3.912	5.988
20	Maine	17.560	14.774	8.453	−1.457	3.339	12.874	4.038	3.754	17.967
21	Maryland	7.337	12.413	11.023	−0.663	3.095	9.505	3.875	3.657	0.488
22	Massachusetts	19.729	10.730	11.021	1.657	2.267	12.840	3.104	3.101	49.490
23	Michigan	12.144	15.254	11.221	0.632	1.970	12.597	2.739	2.772	−0.407
24	Minnesota	8.966	10.105	7.918	0.271	2.613	8.391	3.470	3.387	−2.068
25	Mississippi	22.564	26.027	14.711	−1.435	3.937	21.954	4.342	3.920	−6.072
26	Missouri	18.594	17.068	10.371	−0.330	3.430	14.973	4.093	3.785	7.090

Table A.2: Raw Data, OBP, and MSPE Estimates for Poverty Ratio; Part II

Index	State	y_i	x_{1i}	x_{2i}	x_{3i}	$\sqrt{D_i}$	OBP	MPR	PR	JNR
27	Montana	18.702	20.867	9.800	0.267	3.635	17.737	4.204	3.847	−5.862
28	Nebraska	9.959	13.507	7.526	−0.744	2.610	10.561	3.466	3.384	−2.139
29	Nevada	14.254	14.505	14.382	−0.989	2.748	12.616	3.595	3.475	0.169
30	New Hampshire	6.889	8.322	5.417	−0.300	2.559	6.482	3.416	3.348	−1.767
31	New Jersey	10.844	11.966	10.269	0.822	1.997	10.379	2.774	2.805	−0.402
32	New Mexico	30.031	27.904	13.267	1.351	3.423	25.205	4.089	3.782	21.039
33	New York	20.513	18.890	16.535	1.929	1.593	18.664	2.201	2.216	3.750
34	North Carolina	18.524	18.109	12.522	0.503	2.425	16.382	3.278	3.243	2.817
35	North Dakota	14.846	14.278	7.741	1.703	3.038	12.734	3.833	3.631	0.989
36	Ohio	14.071	14.710	8.251	0.123	1.876	12.953	2.613	2.648	0.938
37	Oklahoma	14.440	22.911	14.050	−0.247	3.760	18.651	4.265	3.879	10.536
38	Oregon	12.613	16.377	13.473	−3.050	3.093	12.947	3.874	3.656	−2.745
39	Pennsylvania	11.507	14.905	9.070	0.534	1.796	12.018	2.502	2.535	0.047
40	Rhode Island	14.148	13.930	13.698	0.515	3.454	12.310	4.107	3.793	−2.524
41	South Carolina	15.620	20.397	12.916	1.714	3.887	17.414	4.321	3.909	−5.017
42	South Dakota	6.778	16.443	8.158	2.114	3.810	12.986	4.287	3.891	35.071
43	Tennessee	16.468	19.710	10.882	−0.876	3.306	16.345	4.018	3.742	−5.680
44	Texas	19.968	23.311	13.450	−0.873	1.684	19.768	2.338	2.364	0.051
45	Utah	6.171	12.421	10.474	−2.212	2.675	8.718	3.529	3.429	4.548
46	Vermont	9.623	13.314	7.720	−2.027	3.240	10.164	3.975	3.717	−4.372
47	Virginia	9.521	14.399	11.296	−0.009	3.035	11.450	3.831	3.630	−0.576
48	Washington	8.368	13.637	11.135	−1.273	3.063	10.422	3.852	3.642	0.070
49	West Virginia	21.021	23.316	13.393	−0.386	3.719	19.922	4.245	3.869	−6.334
50	Wisconsin	6.722	11.660	7.395	1.632	2.949	9.104	3.765	3.588	3.120
51	Wyoming	11.314	16.047	7.311	−0.494	2.933	12.626	3.751	3.579	−1.820

Table A.3 present the raw data and analysis results for Subsection 5.2 of LMJ20. It is divided into Part I and Part II due to its length.

Table A.3: Data, OBP, and MSPE Estimates for Incubation Period of Covid-19;

Part I

Age group	Indx	Province	y_i	x_i	I_i	$\sqrt{D_i}$	OBP	MPR	PR	JNR
#1	1	Anhui	1.638	18.462	0	0.199	1.759	0.038	0.037	0.024
	4	GuangDong	2.070	12.794	0	0.123	2.072	0.016	0.016	0.009
	5	Guangxi	2.173	15.200	0	0.321	2.077	0.069	0.061	−0.010
	6	Guizhou	2.303	20.000	0	0.718	1.959	0.110	0.079	−0.216
	7	Hainan	1.426	12.250	0	0.359	1.879	0.077	0.066	0.181
	8	Hebei	2.708	23.000	0	0.718	1.959	0.110	0.079	0.249
	9	Henan	1.718	16.077	0	0.199	1.828	0.038	0.037	0.022
	11	Hunan	2.629	14.000	0	0.415	2.202	0.086	0.071	0.137
	13	Jiangsu	1.609	20.000	0	0.718	1.885	0.110	0.079	−0.270
	14	Jiangxi	1.099	23.000	0	0.718	1.787	0.110	0.079	0.139
	16	Neimeng	0.693	22.000	1	0.718	0.538	0.110	0.079	0.037
	18	Qinghai	2.485	10.000	0	0.718	2.194	0.110	0.079	−0.155
	19	Shandong	1.861	17.000	0	0.293	1.929	0.063	0.057	−0.006
	21	Shanxi2	0.948	16.333	0	0.239	1.452	0.050	0.047	0.270
	22	Sichuang	1.944	16.812	0	0.180	1.957	0.033	0.032	0.011
	25	Yunan	1.778	10.000	0	0.508	2.086	0.097	0.075	−0.009
	26	Zhejiang	1.968	15.241	0	0.165	1.983	0.028	0.027	0.011
	27	Chongqing	2.322	2.792	0	0.508	2.390	0.097	0.075	−0.015
#2	1	Anhui	1.675	40.964	0	0.079	1.686	0.007	0.007	0.005
	2	Peking	0.693	42.000	0	0.718	1.684	0.110	0.079	0.695
	3	Gansu	1.800	32.000	0	0.321	1.790	0.069	0.061	−0.020
	4	GuangDong	1.681	38.986	0	0.059	1.686	0.004	0.004	0.003
	5	Guangxi	1.918	37.292	0	0.147	1.882	0.023	0.022	0.012
	6	Guizhou	2.773	27.000	0	0.718	1.921	0.110	0.079	0.425
	7	Hainan	1.571	38.500	0	0.147	1.626	0.023	0.022	0.014
	8	Hebei	1.673	39.417	0	0.207	1.720	0.041	0.039	0.011
	9	Henan	1.830	40.444	0	0.060	1.828	0.004	0.004	0.003
	10	Heilongjiang	1.792	32.000	0	0.718	1.785	0.110	0.079	−0.324
	11	Hunan	1.793	41.419	0	0.084	1.794	0.007	0.007	0.006
	12	Jilin	1.657	40.333	0	0.415	1.756	0.086	0.071	−0.054
	13	Jiangsu	1.911	39.842	0	0.117	1.889	0.015	0.015	0.009

Table A.3: Data, OBP, and MSPE Estimates for Incubation Period of Covid-19;

Part II

Age group	Indx	Province	y_i	x_i	I_i	$\sqrt{D_i}$	OBP	MPR	PR	JNR
	14	Jiangxi	1.792	37.000	0	0.718	1.780	0.110	0.079	-0.314
	15	Liaoning	1.806	35.214	0	0.192	1.796	0.036	0.035	0.010
	16	Neimeng	0.936	37.000	1	0.321	0.609	0.069	0.061	0.115
	17	Ningxia	0.645	30.667	0	0.293	1.313	0.063	0.057	0.464
	18	Qinghai	2.129	37.167	0	0.293	1.925	0.063	0.057	0.035
	19	Shandong	1.800	39.737	0	0.082	1.799	0.007	0.007	0.005
	20	Shanxi1	2.226	37.125	0	0.254	1.997	0.054	0.050	0.057
	21	Shanxi2	1.342	40.567	0	0.093	1.397	0.009	0.009	0.010
	22	Sichuang	1.982	39.535	0	0.077	1.965	0.006	0.006	0.005
	23	Tianjin	1.679	39.824	0	0.174	1.715	0.031	0.030	0.012
	24	Xinjiang	0.000	49.500	1	0.508	0.426	0.097	0.075	0.190
	25	Yunan	1.918	34.333	0	0.415	1.815	0.086	0.071	-0.053
	26	Zhejiang	1.551	39.815	0	0.055	1.562	0.003	0.003	0.003
	27	Chongqing	1.861	42.111	0	0.138	1.847	0.021	0.020	0.010
#3	1	Anhui	1.772	62.833	0	0.147	1.887	0.023	0.022	0.024
	3	Gansu	2.639	57.000	0	0.718	2.113	0.110	0.079	0.006
	4	GuangDong	1.643	63.079	0	0.082	1.701	0.007	0.007	0.009
	5	Guangxi	1.397	62.375	0	0.254	1.809	0.054	0.050	0.177
	7	Hainan	1.875	64.714	0	0.192	2.027	0.036	0.035	0.033
	8	Hebei	1.987	64.417	0	0.207	2.103	0.041	0.039	0.022
	9	Henan	1.715	63.519	0	0.138	1.840	0.021	0.020	0.026
	11	Hunan	2.169	64.737	0	0.165	2.203	0.028	0.027	0.012
	13	Jiangsu	2.073	63.692	0	0.199	2.141	0.038	0.037	0.014
	14	Jiangxi	3.045	64.000	0	0.718	2.340	0.110	0.079	0.316
	15	Liaoning	2.773	65.000	0	0.718	2.342	0.110	0.079	-0.010
	19	Shandong	2.280	65.273	0	0.217	2.288	0.043	0.041	0.008
	20	Shanxi1	2.663	76.000	0	0.293	2.710	0.063	0.057	0.012
	21	Shanxi2	1.551	66.400	0	0.185	1.834	0.034	0.033	0.092
	22	Sichuang	1.895	65.000	0	0.185	2.037	0.034	0.033	0.031
	23	Tianjin	2.230	62.889	0	0.239	2.225	0.050	0.047	0.005
	25	Yunan	1.878	66.667	0	0.415	2.225	0.086	0.071	0.069
	26	Zhejiang	1.928	64.018	0	0.095	1.970	0.010	0.010	0.009
	27	Chongqing	1.872	69.083	0	0.207	2.106	0.041	0.039	0.065

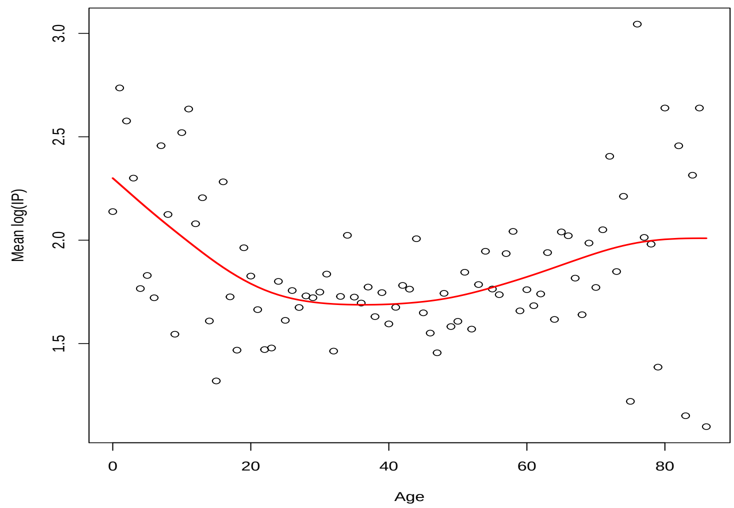
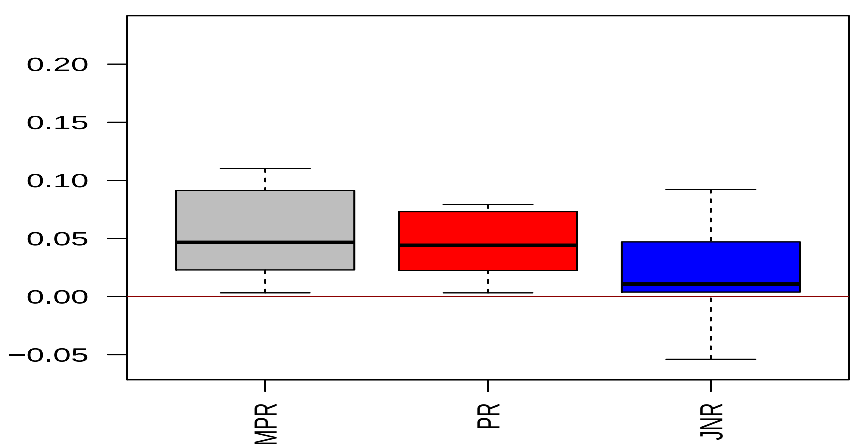
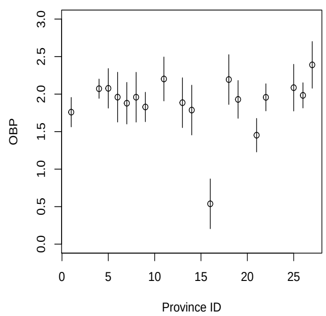


Figure A.1: Scatter Plot of Mean $\log(\mathrm{IP})$ against Age

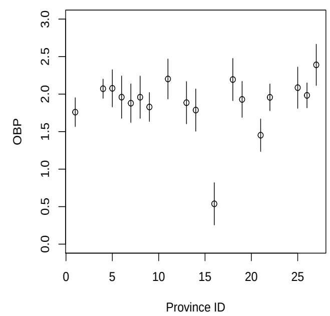


(a) Quadratic model + Indicator.

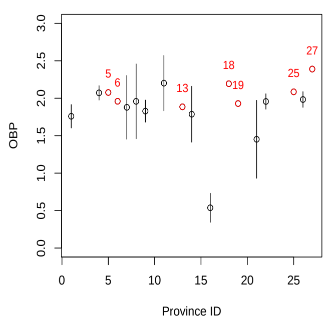
Figure A.2: Boxplots of MSPE estimates



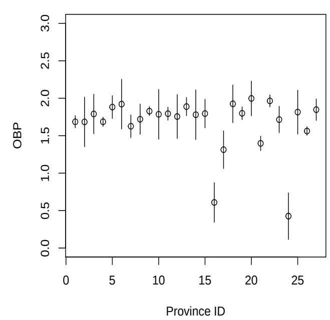
(b) MPR, Age group #1



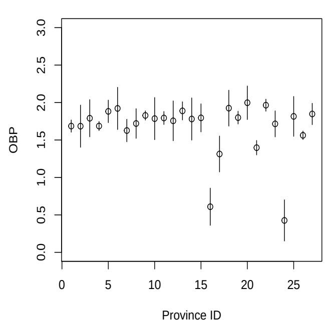
(c) PR, Age group #1



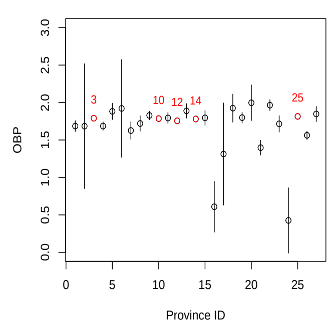
(d) JNR, Age group #1



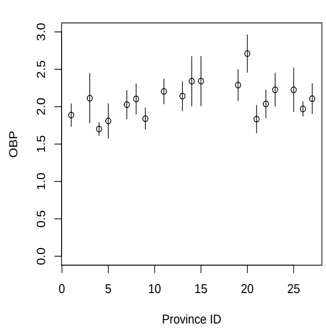
(e) MPR, Age group #2

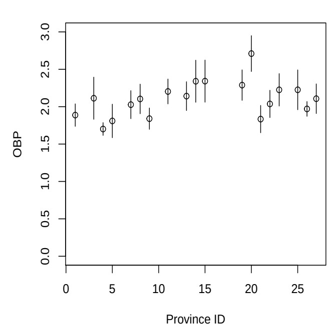


(f) PR, Age group #2



(g) JNR, Age group #2





S4 Small area income and poverty estimation

The Small Area Income and Poverty Estimation (SAIPE) provides estimates of poverty measures for various age groups at the state and county levels of the United States. One measure of particular interest is poverty ratio for school-age children from 5 to 17-year old, used by the U.S. Department of Education to implement its No Child Left Behind program. Following Datta & Mandal (2015), we consider the state level data based on the 1999 Current Population Survey (CPS). The direct estimate y_i for the ith state is computed from the CPS. The estimate is usually subject to large sampling error due to the small sample size. Three covariates are used to capture at least part of the variation, namely, x_1 represents the Internal Revenue Service (IRS) data measuring poverty ratio based on the number of child exemptions; x_2 is the IRS non-filer rate; and x_3 is the residual obtained by fitting a model for the 1989 Census poverty data on x_1, x_2. Let x_{1i}, x_{2i}, x_{3i} denote the observed values of x_1, x_2, x_3, respectively, for the ith state. The sampling variances, D_i, are available from the data. See Table A.2 for detail.

Figure A.4 provides the scatter plots of the direct estimate y's versus the covariates. The dash lines correspond to smoothing spline; solid lines to simple linear regression; a quadratic curve is also fitted versus x_2. It is seen that the marginal relationships between y and x_1 and x_3 are roughly linear, while there might be some nonlinearity in the relationship between y and x_2. Instead of trying to model the nonlinear relationship, a practitioner may prefer to keep the model relatively simple, such as the following type of Fay-Herriot model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + v_i + e_i, \quad i = 1, 2, \cdots, 51. \qquad \text{(S4.1)}$$

On the other hand, due to the potential nonlinearity, the assumed model, (S4.1), may subject to misspecification of the mean function. It is therefore natural to consider OBP , which is more robust to model misspecification than EBLUP.

The BPE of the model parameters are $\hat{\beta}_0 = -1.114$, $\hat{\beta}_1 = 0.860$, $\hat{\beta}_2 = 0.066$, $\hat{\beta}_3 = 0.266$, and $\hat{A} = 2.658$. We then obtain the OBPs and their corresponding MSPE estimates using the MPR, PR and JNR methods for the 51 U.S. states. The results are reported in Table A.2, along with the original data. It is observed that, for the JNR MSPE estimates, 21 out of 51, or about 41%, of the values are negative. On the other hand, all of the MPR and PR estimates are positive.

Figure A.5 shows boxplots of the MSPE estimates. Figure A.6 plots the OBP vs the state index with the square root of the corresponding MSPE estimate (plus/minus) used as a margin of error. Note that many of these margins of error are not available for JNR due to the negative MSPE estimates.

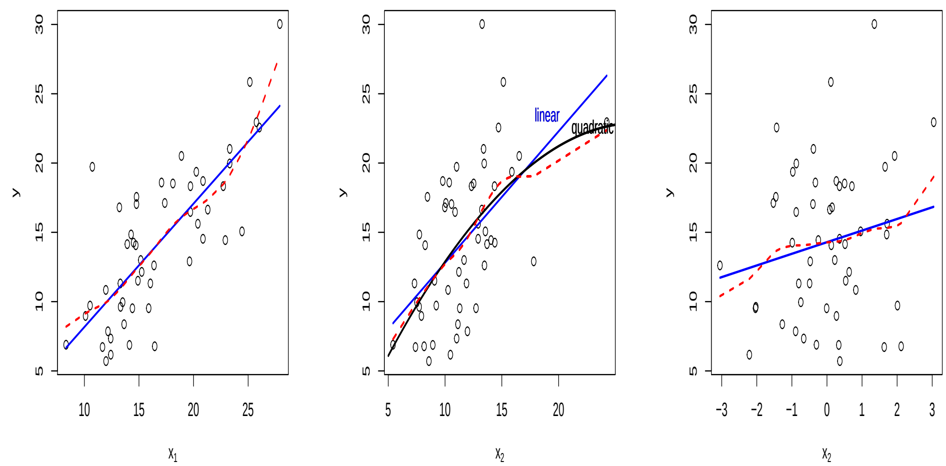


Figure A.4: Direct Estimate vs Covariates

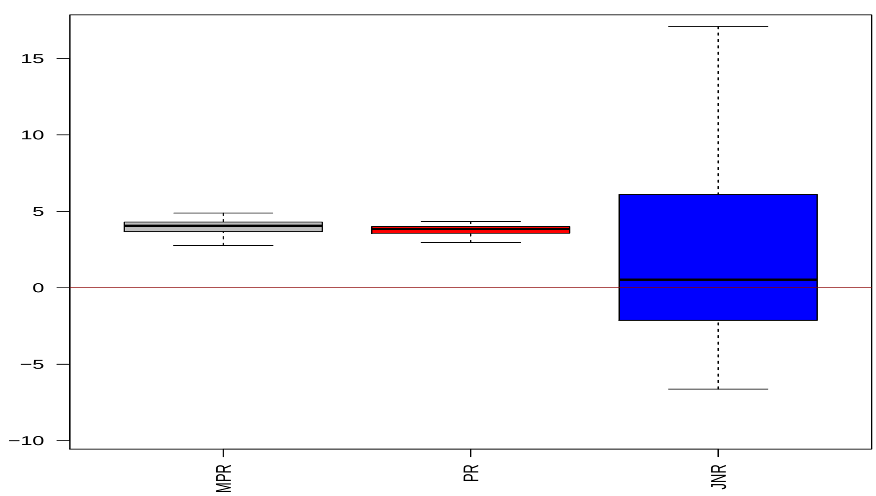
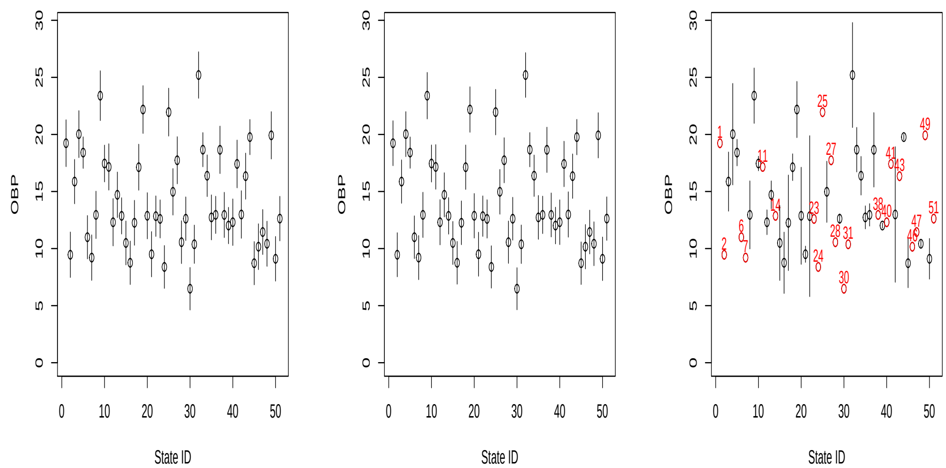


Figure A.5: Boxplots of MSPE estimates for the 51 States



(n) (o) (p)

Figure A.6: OBP with Margin of Error. Length Line Equal to 2 Times Square Root of Corresponding MSPE Estimate; (a) MPR; (b) PR; (c) JNR; Dot with No Line Indicating Negative MSPE Estimates (Number Indicating Corresponding State)

Bibliography

Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer, New York.