Supplementary Material to Sumca: Simple, Unified, Monte-Carlo Assisted Approach to Second-order Unbiased MSPE Estimation

Jiming Jiang[1] and Mahmoud Torabi[2]
[1] *University of California, Davis and* [2] *University of Manitoba*

A.1 Technical results and proofs

Following Jiang *et al.* (2002) and Das, Jiang and Rao (2004), we consider a sequence of subsets, $\Psi_m \subset \Psi$, where Ψ is the natural parameter space for ψ, and Ψ_m is a compact subset that lies strictly in Ψ^o, the interior of Ψ, and approaches Ψ as $m \to \infty$ in the sense that $\Psi_m \subset \Psi_{m+1}$ for any $m \geq 1$, and any point in Ψ^o will be covered by Ψ_m for sufficiently large m. For example, for the Fay-Herriot model, the parameter vector is $\psi = (\beta', A)'$, whose natural parameter space is $\Psi = \{\psi : \beta \in R^p, A \geq 0\}$, where R^p is the p-dimensional Euclidean space. A choice of Ψ_m may be $\Psi_m = \{\psi : |\beta| \leq C(\log m)^L, c_1(\log m)^{-L} \leq A \leq c_2(\log m)^L\}$, where C, L, c_1, c_2 are any fixed positive constants. Let $\hat{\psi}_{\rm o}$ be the original estimator of ψ. A truncated estimator (e.g., Das *et al.* 2004) is defined as $\hat{\psi} = \hat{\psi}_{\rm o}$ if $\hat{\psi}_{\rm o} \in \Psi_m$, and $\hat{\psi} = \psi_*$ otherwise, where ψ_* is a known vector that belongs to Ψ_m. Note that, if the true $\psi \in \Psi^o$ and $\hat{\psi}_{\rm o}$ is a consistent estimator, then we have with probability tending to one that $\hat{\psi} = \hat{\psi}_{\rm o}$; in other words, asymptotically, the truncated estimator is equal to the original estimator. Also, since the constants C, L, c_2 can be chosen arbitrarily large, and c_1 arbitrarily close to 0, in practice the truncation does not change the value of the estimator.

1. Notations and a lemma. We first state a lemma that plays a key role in verifying a key assumption of Theorem 1. The lemma is regarding an asymptotic expansion of an estimator that is a root to an estimating equation. Expansion of this type has a well-traveled line in the literature, for example, in the context of maximum likelihood estimation (e.g., Cox and Hinkley 1973, ch.9, Pfanzagl 1980, Das *et al.* 2004, and Jiang 2010, ch. 4). The proof of the lemma is deferred to an online manuscript in arxiv (Jiang and Torabi 2019).

Consider an objective function of the following form, $l(\psi) = \sum_{t=1}^m l_t(\psi, z_t)$, where $z_t, 1 \leq t \leq m$ are independent random vectors. Define $J_\psi = (\partial/\partial\psi)l(\psi)$, $H_\psi = {\rm E}_\psi\{(\partial^2/\partial\psi\partial\psi')l(\psi)\}$, $D_\psi = (\partial^2/\partial\psi\partial\psi')l(\psi) - H_\psi$, $q_{j,\psi} = (\partial^3/\partial\psi\partial\psi'\partial\psi_j)l(\psi)$, where ψ_j is the jth component of ψ, and $Q_{j,\psi} = {\rm E}_\psi(q_{j,\psi})$. Here, ${\rm E}_\psi$ and ${\rm P}_\psi$ mean expectation and probability, respectively, under the distribution with ψ being the true parameter vector. Let $g(\psi) = \partial l/\partial\psi = \sum_{t=1}^m g_t(\psi, z_t)$ be an estimating function, where $g_t(\psi, z_t) = \partial l_t/\partial\psi$. The estimator, $\hat{\psi}$, is a root to the estimating equation $g(\psi) = 0$.

It should be noted that, in asymptotic analysis of mixed effects models (e.g., Miller

1977, Jiang 1996), various notations such as the X and D in Section 4.2, and those involved in (A.3) below, depend on the sample sizes, which are (much) more complicated than in the i.i.d. case. For example, in the case of Fay-Herriot model (Section 4.1), the sample size is determined by m; in the case of mixed logistic model (Section 4.3), the sample sizes depend not only on m but also on $n_i, 1 \le i \le m$. However, following the standard notational treatment as in the above references, the dependency on the sample size is suppressed from the notation. So, for example, we will write X instead of X_m in Section 4.2. The same simplification also applies to the notations involved in (A.3) below and elsewhere.

Lemma A.1. Suppose that (i) l_t is four-times continuously differentiable with respect to ψ, $1 \le t \le m$; (ii) there is a constant $0 < \delta < 1/24$ such that

$$\liminf_{m\to\infty} m^{\delta-1} \inf_{\psi\in\Psi_m} \lambda_{\min}(H_\psi) > 0; \tag{A.1}$$

(iii) there is $M \ge 1$ such that when $m \ge M$ we have $S_{m,\psi} = \{\tilde{\psi} : |\tilde{\psi} - \psi| \le m^{4\delta-1/2}\} \subset \Psi^o$, the interior of Ψ, for all $\psi \in \Psi_m$; (iv) $\mathrm{E}_\psi\{(\partial/\partial\psi)l_t(\psi, z_t)\} = 0$, $1 \le t \le m$ and there is a constant $b \ge 2$ such that the bth moments of the following are uniformly bounded for $m \ge M$, $\psi \in \Psi_m$, $1 \le t \le m$, and $1 \le j, k, l, q \le s$:

$$m^{-\delta}\left|\frac{\partial}{\partial\psi_j} l_t(\psi, z_t)\right|, m^{-\delta}\left|\frac{\partial^2}{\partial\psi_j\partial\psi_k} l_t(\psi, z_t)\right|, m^{-\delta}\left|\frac{\partial^3}{\partial\psi_j\partial\psi_k\partial\psi_l} l_t(\psi, z_t)\right|,$$

$$m^{-\delta}\sup_{\tilde{\psi}\in S_{m,\psi}}\left|\frac{\partial^4}{\partial\psi_j\partial\psi_k\partial\psi_l\partial\psi_q} l_t(\tilde{\psi}, z_t)\right|, \tag{A.2}$$

Then, there is $N \ge 1$ such that the following holds for $m \ge N$:
(I) for every $\psi \in \Psi_m$ there is a $\hat{\psi} \in \Psi_m$ and event set $\mathcal{B}_{m,\psi}$ such that on $\mathcal{B}_{m,\psi}$ we have $(\partial/\partial\psi)l(\hat{\psi}) = 0$ and $|\hat{\psi} - \psi| \le c_1 m^{3\delta-1/2}$, where c_1 is a constant not depending on ψ;
(II) $\mathrm{P}_\psi(\mathcal{B}^c_{m,\psi}) \le c_2 m^{-b\delta}$, where c_2 is a constant not depending on ψ;
(III) on $\mathcal{B}_{m,\psi}$, the following asymptotic expansion holds for $m \ge M$:

$$\hat{\psi} - \psi = -H_\psi^{-1}J_\psi + H_\psi^{-1}D_\psi H_\psi^{-1}J_\psi - \frac{1}{2}H_\psi^{-1}[J'_\psi H_\psi^{-1}Q_{j,\psi}H_\psi^{-1}J_\psi]_{1\le j\le s} + R, \tag{A.3}$$

where $|R| \le m^{12\delta-3/2}u = o(m^{-1})u$ and $\mathrm{E}(u^{b/3})$ is uniformly bounded for $\psi \in \Psi_m$. To see the orders of the terms in the expansion (A.3), consider, for example, the Fay-Herriot model. Then, under regularity conditions, we have $J_\psi = O(m^{1/2})$, $H_\psi = O(m)$, $D_\psi = O(m^{1/2})$ and $Q_{j,\psi} = O(m)$ (note that here ψ stands for the true parameter vector). It follows that the leading term on the right side of (A.3) is $O(m^{-1/2})$, the nest two terms are both $O(m^{-1})$, and R is $o(m^{-1})$.

More notation: Let y_i denote the vector of observations from the ith small area, $i = 1, \dots, m$. The mixed effect of interest is $\theta = \theta_i$, which is a characteristic of interest associated with the ith small area. For example, under the Fay-Herriot model, y_i is the direct survey estimator from the ith small area, and θ_i is the small area mean, which can be expressed as $\theta_i = x_i'\beta + v_i$. Under the mixed logistic model, $y_i = (y_{ij})_{1\leq j\leq n_i}$, where $y_{ij}, 1 \leq j \leq n_i$ are binary outcomes sampled from the ith small area, and $\theta_i = p_i$, which is a conditional probability associated with the ith area.

Suppose that, when the true ψ is known, the predictor of θ_i can be expressed as

$$\tilde{\theta}_i = h_0(y_i, \psi) \tag{A.4}$$

for some function $h_0(\cdot, \cdot)$. Assumption *B0* below implies that the conditional moments, $h_j(y, \psi) = \mathrm{E}(\theta_i^j|y), j = 1, 2$ involved in (2) have similar expressions:

$$h_j(y, \psi) = h_j(y_i, \psi), \quad j = 1, 2 \tag{A.5}$$

for some functions $h_j(\cdot, \cdot), j = 1, 2$ [because $\mathrm{E}(\theta_i^j|y) = \mathrm{E}(\theta_i^j|y_i)$ (e.g., Shao 2003, p. 41)].

Strictly speaking, the functions $h_j(\cdot), j = 0, 1, 2$ also depend on i. But, because we are considering the MSPE of $\hat{\theta}_i$ for a fixed i, the dependency of i is suppressed from the notation (e.g., h_0 instead of h_{i0}). Likewise, O's and o's that appear in this paper, such as those in Theorems 1–4 and Theorem A.1 below, may depend on i, and therefore non-uniform across different i's. This is standard in the SAE literature (e.g., Rao and Molina 2015, Jiang 2017, ch. 4).

Sometimes, instead of considering y_t directly, we consider $z_t = y_t - \mathrm{E}(y_t)$, $1 \leq t \leq m$, where $\mathrm{E}(y_t)$ is the true expectation of y_t. Of course, z_t is unobserved, but this does not affect our theoretical arguments. In such a case, the right sides of (A.4), (A.5) will be written as $h_j(z_i, \psi), j = 0, 1, 2$, respectively, for (slightly) different functions $h_j, j = 0, 1, 2$. The theory below will be developed under the notation z_t, which includes the case $z_t = y_t$ and the case $z_t = y_t - \mathrm{E}(y_t)$.

It should be noted that an objective function in Lemma A.1 does not always exist; nevertheless, the estimating function, hence the estimating equation, exists more widely (see Section A.2.1 of Jiang and Torabi 2019). What is a key to our proof is an asymptotic expansion like (A.3), whose derivation can be done via the solution to an estimating equation. The objective function is only used in the first half of Lemma A.1 [i.e., (I) and (II)] to ensure the existence of a solution to an estimating equation; however, if a solution to the estimating equation, such as (A.6) below, is known to exist, one does not need an objective function to obtain the asymptotic expansion.

Sometimes, the estimating function may not be expressed as sum of g_t, which only depends on z_t; an additional term, $\Delta(\psi, z)$, is involved, where $z = (z_t)_{1\leq t\leq m}$, so that the estimating equation can be expressed as

$$\sum_{t=1}^{m} g_t(\psi, z_t) + \Delta(\psi, z) = 0 \tag{A.6}$$

(again, see Section A.2.1 of Jiang and Torabi 2019). When dealing directly with the estimating function, the previous notation H_ψ and $Q_{j,\psi}$ are now extended, with $\partial l_t/\partial\psi$ replaced by $g_t(\psi, z_t)$. Note that, in case the objective function exists, we have $\Delta(\psi, z) = 0$.

Define $\bar{H}_\psi = m^{-1}H_\psi$, $\bar{Q}_{j,\psi} = m^{-1}Q_{j,\psi}$, $D_t(\psi, z_t) = \partial^2 l_t/\partial\psi\partial\psi' - \mathrm{E}_\psi(\partial^2 l_t/\partial\psi\partial\psi')$, $p_j(z_i, \psi) = \partial h_j/\partial\psi, j = 0, 1$, $d_1(z_i, \psi) = \partial h_2/\partial\psi - 2h_0(\partial h_1/\partial\psi)$, and $d_2(z_i, \psi) = \partial^2 h_2/\partial\psi\partial\psi' - 2h_0(\partial^2 h_1/\partial\psi\partial\psi')$ [see (A.4), (A.5) for the definitions of $h_j, j = 0, 1, 2$]. The theorem below shows that, when the objective function exists, *A2* holds with

$$q(\psi) = q_1(\psi) - q_2(\psi) + \frac{1}{2}\{q_3(\psi) - q_4(\psi)\} + 2q_5(\psi), \tag{A.7}$$

where $q_1(\psi) = \mathrm{E}_\psi\{\phi_{1,i}(z_i, \psi)\}$ and $q_j(\psi) = m^{-1}\sum_{t=1}^m \mathrm{E}_\psi\{\phi_{j,t}(z_i, z_t, \psi)\}$, $j = 2, 3, 4, 5$ with $\phi_{1,i}(z_i, \psi) = d_1(z_i, \psi)'\bar{H}_\psi^{-1}g_i(\psi, z_i)$,

$$\begin{aligned}
\phi_{2,t}(z_i, z_t, \psi) &= d_1(z_i, \psi)'\bar{H}_\psi^{-1}D_t(\psi, z_t)\bar{H}_\psi^{-1}g_t(\psi, z_t), \\
\phi_{3,t}(z_i, z_t, \psi) &= d_1(z_i, \psi)'\bar{H}_\psi^{-1}[g_t(\psi, z_t)'\bar{H}_\psi^{-1}\bar{Q}_{j,\psi}\bar{H}_\psi^{-1}g_t(\psi, z_t)]_{1\le j\le s}, \\
\phi_{4,t}(z_i, z_t, \psi) &= g_t(\psi, z_t)'\bar{H}_\psi^{-1}d_2(z_i, \psi)\bar{H}_\psi^{-1}g_t(\psi, z_t), \\
\phi_{5,t}(z_i, z_t, \psi) &= g_t(\psi, z_t)'\bar{H}_\psi^{-1}p_0(z_i, \psi)p_1(z_i, \psi)'\bar{H}_\psi^{-1}g_t(\psi, z_t).
\end{aligned}$$

When the objective function does not exist, but (ii) of *B2* below holds, *A2* holds with

$$q(\psi) = q_1(\psi) - q_2(\psi) + \frac{1}{2}\{q_3(\psi) - q_4(\psi)\} + 2q_5(\psi) + q_6(\psi), \tag{A.8}$$

where $q_6(\psi) = \mathrm{E}_\psi\{\phi_{6,i}(z, \psi)\}$ with $\phi_{6,i}(z, \psi) = d_1(z_i, \psi)'\bar{H}_\psi^{-1}\Delta(\psi, z)$.

2. Verifiable conditions for Theorem 1. Assumption *A1* of Theorem 1 is a natural condition. On the other hand, assumption *A2* may not be straightforward to verify. Here, we focus on the SAE case, and provide sufficient conditions for *A2* that are easier to verify.

Theorem A.1. The assumptions of Theorem 1, hence its conclusion, hold under (A.4), (A.5) and conditions *B0–B5* below for some $0 < \delta < 1/28$ and any $b > 0$:

B0. $(\theta_i, y_i), i = 1, \dots, m$ are independent,

B1. There is $M \ge 1$ such that when $m \ge M$, the true $\psi \in \Psi_m$, and $\psi \in \Psi_m$ implies $|\psi| \le C(\log m)^L$ for some constants $C, L > 0$ and $S_{m,\psi} = \{\tilde{\psi} : |\tilde{\psi} - \psi| \le m^{4\delta-1/2}\} \subset \Psi^{\mathrm{o}}$.

B2. Either (i) the objective function exists and $\Delta(\psi, z) = 0$; or (ii) $\hat{\psi}$ satisfies (A.6), H_ψ is symmetric and, for any $\psi \in \Psi_m$, there is $\mathcal{B}_{m,\psi}$ such that $|\hat{\psi} - \psi| \le cm^{3\delta-1/2}$ on $\mathcal{B}_{m,\psi}$ and $\mathrm{P}_\psi(\mathcal{B}_{m,\psi}^c) \le c_b m^{-b\delta}$, where c, c_b are constants that do not depend on ψ.

B3. $\mathrm{E}_\psi\{g_t(\psi, z_t)\} = 0, 1 \le t \le m$, (A.1) holds, and $\mathrm{E}(\theta_i^2) < \infty$.

B4. There are constants $c_j > 0, K_j \ge 0$ and functions $f_j(\cdot)$, $j = 0, 1, 2$ such that the absolute values of the following and their up to 3rd-order partial derivatives, with respect to ψ: (i) $g_t(\psi, z_t)$, $1 \le t \le m$, (ii) $h_j(z_i, \psi)$, $j = 0, 1, 2$, and (iii) $\Delta(\psi, z)$ are bounded by (i) $c_1(|\psi| \vee 1)^{K_1}f_1(z_t)$, $1 \le t \le m$, (ii) $c_2(|\psi| \vee 1)^{K_2}f_2(z_i)$, and (iii) $c_0(|\psi| \vee 1)^{K_0}f_0(z)$,

respectively. Furthermore, there is $0 < \delta_1 < \delta$ such that $m^{-\delta_1 b}\sup_{\psi\in\Psi_m}\mathrm{E}_\psi\{f_1^b(z_t)\}$, $1 \le t \le m$, $m^{-\delta_1 b}\sup_{\psi\in\Psi_m}\mathrm{E}_\psi\{f_2^b(z_i)\}$, and $m^{-\delta_1 b}\sup_{\psi\in\Psi_m}\mathrm{E}_\psi\{f_0^b(z)\}$ are bounded.
B5. We have the following expressions: $\mathrm{E}_\psi\{\phi_{1,i}(z_i,\psi)\} = \mathrm{E}\{w_{1,i}(\xi_i,\psi)\}$,

$$\mathrm{E}_\psi\{\phi_{j,t}(z_i,z_t,\psi)\} = \mathrm{E}\{w_{j,t}(\xi_i,\xi_t,\psi)\}, j = 2,3,4,5, \mathrm{E}_\psi\{\phi_{6,i}(z,\psi)\} = \mathrm{E}\{w_{6,i}(\xi,\psi)\},$$

where $z = (z_t)_{1\le t\le m}, \xi = (\xi_t)_{1\le t\le m}$ such that the distribution of ξ does not depend on ψ. Furthermore, there are constants $c_3 > 0, K_3 \ge 0$ and functions $f_3(\cdot)$, $f_4(\cdot,\cdot)$, $f_5(\cdot)$ such that the absolute values of the first-order partial derivatives of the following, with respect to ψ: (i) $w_{1,i}(\xi_i,\psi)$, (ii) $w_{j,t}(\xi_i,\xi_t,\psi)$, $j = 2,3,4,5$, and (iii) $w_{6,i}(\xi,\psi)$ are bounded by (i) $c_3(|\psi|\vee 1)^{K_3} f_3(\xi_i)$, (ii) $c_3(|\psi|\vee 1)^{K_3} f_4(\xi_i,\xi_t)$, and (iii) $c_3(|\psi|\vee 1)^{K_3} f_5(\xi)$, respectively, and $\mathrm{E}\{f_3^2(\xi_i)\}$, $\mathrm{E}\{f_4^2(\xi_i,\xi_t)\}$, $1 \le t \le m$, $\mathrm{E}\{f_5^2(\xi)\}$ are bounded.

Proof: We need to verify assumptions *A1*, *A2* of Theorem 1, where ψ stands for the true parameter vector. Without loss of generality, we can assume that $b \ge 2$. Below c denotes a generic constant whose value may be different at different places.

A1: B3 implies that $\mathrm{E}(\theta^2|y) = \mathrm{E}(\theta_i^2|y) < \infty$ with probability one. Also, by (2), (A.4), (A.5), and *B4*, it is easy to show that $|a(y,\psi)| \vee |a(y,\hat{\psi})| \le c(1+\log m)^{2K_2}\{f_2(z_i)\vee 1\}^2$. Thus, both $\mathrm{E}\{a(y,\psi)\}$ and $\mathrm{E}\{a(y,\hat{\psi})\}$ exist by *B4*.

A2: We first show that the conditions of Lemma A.1 are satisfied in case of (i) of *B2*. Conditions (i)–(iii) are obvious. For (iv), the first part is obvious. As for the boundedness of the bth moments, (A.2), note that $\partial l_t/\partial\psi = g_t(\psi, z_t), 1 \le t \le m$, hence

$$\frac{\partial}{\partial\psi_j} l_t(\psi,z_t) = g_{tj}(\psi,z_t), \cdots, \frac{\partial^4}{\partial\psi_j\partial\psi_k\partial\psi_l\partial\psi_q} l_t(\tilde{\psi},z_t) = \frac{\partial^3}{\partial\psi_k\partial\psi_l\partial\psi_q} g_{tj}(\tilde{\psi},z_t),$$

where $g_{tj}(\cdot,\cdot)$ is the jth component of $g_t(\cdot,\cdot)$. Thus, we have, for any $\psi\in\Psi_m$,

$$\left|\frac{\partial}{\partial\psi_j} l_t(\psi,z_t)\right| = |g_{tj}(\psi,z_t)| \le c_1(|\psi|\vee 1)^{K_1} f_1(z_t) \le c(1+\log m)^{LK_1} f_1(z_t),$$

$$\mathrm{E}\left(m^{-\delta}\left|\frac{\partial}{\partial\psi_j} l_t(\psi,z_t)\right|\right)^b \le cm^{-\delta b}(1+\log m)^{LK_1 b}\mathrm{E}\{f_1^b(z_t)\} \le cm^{-\delta_1 b}\mathrm{E}\{f_1^b(z_t)\},$$

which is bounded. Similarly, one can show that the bth moments of $m^{-\delta}|(\partial/\partial\psi_k)g_{tj}(\psi,z_t)|$ and $m^{-\delta}|(\partial^2/\partial\psi_k\partial\psi_l)g_{tj}(\psi,z_t)|$ are bounded for $\psi\in\Psi_m$. Finally, since $\psi\in\Psi_m$ and $\tilde{\psi}\in S_{m,\psi}$ imply $|\tilde{\psi}| \le |\psi| + m^{4\delta-1/2} \le C(\log m)^L + m^{4\delta-1/2} \le c(1+\log m)^L, m \ge M$, by a similar argument, it can be shown that the bth moments of

$$m^{-\delta}\sup_{\tilde{\psi}\in S_{m,\psi}}\left|\frac{\partial^3}{\partial\psi_k\partial\psi_l\partial\psi_q} g_{tj}(\tilde{\psi},z_t)\right|$$

are bounded. Thus, all of the conditions of Lemma A.1 are verified under case (i) of *B2*.

Next, we show that, under case (ii) of *B2*, the conclusion of part (III) of Lemma A.1 holds. This follows by going through the proof of the lemma (see Jiang and Torabi 2019). In particular, (A.1) continues to hold because, with $J_4 = \Delta(\hat{\psi}, z)$, we have $|H_\psi^{-1} J_4| \le cm^{\delta-1}(1+\log m)^{LK_0} f_0(z) \le cm^{3\delta-1} u_4$ and $\mathrm{E}_\psi(u_4^b)$ uniformly bounded for $\psi \in \Psi_m$. This leads to the next-step expansion, the one with $R_1 + R_2 + R_3$. Now two more terms will be added to this equation: $\Delta(\psi, z)$ and $R_4 = \{(\partial/\partial\psi')\Delta(\tilde{\psi}, z)\}(\hat{\psi} - \psi)$, where $\tilde{\psi}$ lies between ψ and $\hat{\psi}$. By *B3*, *B4*, it is easy to show that $|H_\psi^{-1} R_4| \le cm^{5\delta-3/2} U_7$ with $\mathrm{E}(U_7^b)$ uniformly bounded for $\psi \in \Psi_m$. It follows that the asymptotic expansion (A.3) holds with an additional term, $-H_\psi^{-1}\Delta(\psi, z)$, on the right side before R, whose order is unchanged.

Now, in view of Lemma A. 1, we can write

$$d(\psi) = \mathrm{E}_\psi[\{a(y,\psi) - a(y,\hat{\psi})\}1_{\mathcal{B}}] + \mathrm{E}_\psi[\{a(y,\psi) - a(y,\hat{\psi})\}1_{\mathcal{B}^c}] = I_1 + I_2, \qquad \text{(A.9)}$$

where $\mathcal{B}$ is the $\mathcal{B}_{m,\psi}$ in Lemma A.1. Note that

$$a(y,\psi) - a(y,\hat{\psi}) = 2h_0(z_i,\hat{\psi})\{h_1(z_i,\hat{\psi}) - h_1(z_i,\psi)\} - \{h_2(z_i,\hat{\psi}) - h_2(z_i,\psi)\}. \qquad \text{(A.10)}$$

Thus, by *B4*, it is easy to show that $|a(y,\psi) - a(y,\hat{\psi})| \le c(1+\log m)^{2LK_2}\{f_2(z_i) \vee 1\}^2$. It follows, by the Cauchy-Schwarz inequality, that

$$I_2 \le [\mathrm{E}\{|a(y,\psi) - a(y,\hat{\psi})|^2\}]^{1/2} \mathrm{P}^{1/2}(\mathcal{B}^c) \le c(1+\log m)^{2LK_2} m^{-b(\delta-\delta_1)/2}. \qquad \text{(A.11)}$$

The right side of (A.11) is $o(m^{-1})$ uniformly for $\psi \in \Psi_m$ if $b > 4 \vee \{2(\delta - \delta_1)^{-1}\}$. We next consider I_1. First consider case (i) of *B2*. From (A.10), we can further write

$$\begin{aligned} a(y,\psi) - a(y,\hat{\psi}) \;=\; & 2h_0(z_i,\psi)\{h_1(z_i,\hat{\psi}) - h_1(z_i,\psi)\} - \{h_2(z_i,\hat{\psi}) - h_2(z_i,\psi)\} \\ & +2\{h_0(z_i,\hat{\psi}) - h_0(z_i,\psi)\}\{h_1(z_i,\hat{\psi}) - h_1(z_i,\psi)\}. \qquad \text{(A.12)} \end{aligned}$$

Denote the first three terms on the right side of (A.3) by J_1, J_2, and $-J_3/2$, respectively. Some tedious evaluations, using *B1–B4*, show that

$$\begin{aligned} & \mathrm{E}_\psi\left[h_0^{2-j}\{h_j(z_i,\hat{\psi}) - h_j(z_i,\psi)\}1_{\mathcal{B}}\right] = -\mathrm{E}_\psi\left(h_0^{2-j}\frac{\partial h_j}{\partial\psi'}J_1\right) + \mathrm{E}_\psi\left(h_0^{2-j}\frac{\partial h_j}{\partial\psi'}J_2\right) \\ & -\frac{1}{2}\left\{\mathrm{E}_\psi\left(h_0^{2-j}\frac{\partial h_j}{\partial\psi'}J_3\right) - \mathrm{E}_\psi\left(h_0^{2-j}J_1'\frac{\partial^2 h_j}{\partial\psi\partial\psi'}J_1\right)\right\} + \Delta_j, \quad j = 1,2, \\ & \mathrm{E}_\psi\left[\{h_0(y_i,\hat{\psi}) - h_0(y_i,\psi)\}\{h_1(y_i,\hat{\psi}) - h_1(y_i,\psi)\}1_{\mathcal{B}}\right] = \mathrm{E}_\psi\left(\frac{\partial h_0}{\partial\psi'}J_1\frac{\partial h_1}{\partial\psi'}J_1\right) + \Delta_3 \\ & \text{with } \sup_{\psi\in\Psi_m}|\Delta_j| = o(m^{-1}),\ j = 1,2,3. \qquad \text{(A.13)} \end{aligned}$$

A "trick" that is repeatedly used in the evaluation is the following: Suppose that $J = \sum_{t=1}^m \eta_t(\psi, z_t)$ such that $|\eta_t(\psi, z_t)| \le c(|\psi| \vee 1)^a f(z_t)$ for some constants $c, a > 0$ and function $f(\cdot)$, and $m^{-\delta_1 b}\mathrm{E}_\psi\{f^b(z_t)\}$ are bounded uniformly for $\psi \in \Psi_m$ for some $b \geq 2$. Then the following hold: (i) If $\mathrm{E}_\psi\{\eta_t(\psi, z_t)\} = 0, 1 \leq t \leq m$, the bth moment of $m^{-1/2-\delta}|J|$ is bounded uniformly for $\psi \in \Psi_m$ (this is shown by Burkholder's and Jensen's inequalities; see the proof of Lemma A.1). (ii) Without assuming that the means of η_t are zero, the bth moment of $m^{-1-\delta}|J|$ is bounded uniformly for $\psi \in \Psi_m$ (this is shown by Jensen's inequality). Combining (A.9), (A.11)–(A.13), we obtain

$$\begin{aligned} d(\psi) &= \mathrm{E}_\psi\{d_1'(z_i,\psi)J_1\} - \mathrm{E}_\psi\{d_1'(z_i,\psi)J_2\} + \frac{1}{2}[\mathrm{E}_\psi\{d_1'(z_i,\psi)J_3\} \\ &\quad -\mathrm{E}_\psi\{J_1' d_2(z_i,\psi)J_1\}] + 2\mathrm{E}_\psi\{p_0'(z_i,\psi)J_1 p_1'(z_i,\psi)J_1\} + r(\psi) \end{aligned} \tag{A.14}$$

with $\sup_{\psi\in\Psi_m}|r(\psi)| = o(m^{-1})$.

Furthermore, direct computation shows that $\mathrm{E}_\psi\{d_1'(z_i,\psi)J_j\} = m^{-1}q_j(\psi)$, $j = 1, 2, 3$, $\mathrm{E}_\psi\{J_1'd_2(z_i,\psi)J_1\} = m^{-1}q_4(\psi)$, and $\mathrm{E}_\psi\{p_0'(z_i,\psi)J_1p_1'(z_i,\psi)J_1\} = m^{-1}q_5(\psi)$. This leads to the expression of $d(\psi)$ in *A2* with $q(\psi)$ given by (A.7).

We now consider I_1 in (A.9) under (ii) of *B2*. As noted above, this results in an extra term on the right side of (A.3), $-H_\psi^{-1}\Delta(\psi, z) \equiv -J_4$. Some careful evaluation shows that this results in, up to a term that is $o(m^{-1})$ uniformly for $\psi \in \Psi_m$, an extra term equal to $\mathrm{E}_\psi\{d_1'(z_i,\psi)J_4\} = m^{-1}q_6(\psi)$. Thus, *A2* holds with $q(\psi)$ given by (A.8).

It remains to show that $\mathrm{E}\{|q(\hat\psi) - q(\psi)|\} = o(1)$. It suffices to show that $\mathrm{E}\{|q_j(\hat\psi - q_j(\psi)|\} = o(1)$, $j = 1, \ldots, 6$. We consider $j = 2$ as an example; the rest can be proved similarly. By *B5*, we have $\mathrm{E}_\psi\{\phi_{2,t}(z_i, z_t, \psi)\} = \mathrm{E}\{w_{2,t}(\xi_i, \xi_t, \psi)\}$, so $q_2(\psi) = m^{-1}\sum_{t=1}^m \mathrm{E}\{w_{2,t}(\xi_i,\xi_t,\psi)\}$. Let $\mathrm{E}\{|q_2(\hat\psi)-q_2(\psi)|\} = \mathrm{E}\{|q_2(\hat\psi)-q_2(\psi)|1_{\mathcal{B}}\}+\mathrm{E}\{|q_2(\hat\psi)-q_2(\psi)|1_{\mathcal{B}^c}\} = \Delta_1+\Delta_2$. For Δ_1, note that for any $1 \leq t \leq m$, we have, by Taylor expansion,

$$w_{2,t}(\xi_i,\xi_t,\hat\psi) - w_{2,t}(\xi_i,\xi_t,\psi) = \left\{\frac{\partial}{\partial\psi'}w_{2,t}(\xi_i,\xi_t,\tilde\psi)\right\}(\hat\psi - \psi),$$

where $\tilde\psi$ lies between ψ and $\hat\psi$. For large m we have $\psi \in \Psi_m$; also $\hat\psi \in \Psi_m$ by definition, hence $\tilde\psi \in \Psi_m$. It follows, by *B5*, that on $\mathcal{B}$

$$|w_{2,t}(\xi_i,\xi_t,\hat\psi) - w_{2,t}(\xi_i,\xi_t,\psi)| \leq c(1+\log m)^{LK_3}m^{3\delta-1/2}f_4(\xi_i,\xi_t), \tag{A.15}$$

$1 \leq t \leq m$. It follows that $\Delta_1 \leq c(1+\log m)^{LK_3}m^{3\delta-1/2}m^{-1}\sum_{t=1}^m \mathrm{E}\{f_4(\xi_i,\xi_t)\} = o(1)$. For Δ_2, we have, similar to (A.15) but without using the property of $\mathcal{B}$,

$$\begin{aligned} |w_{2,t}(\xi_i,\xi_t,\hat\psi) - w_{2,t}(\xi_i,\xi_t,\psi)| &\leq c(1+\log m)^{LK_3}f_4(\xi_i,\xi_t)(|\hat\psi|+|\psi|) \\ &\leq c(1+\log m)^{L(K_3+1)}f_4(\xi_i,\xi_t), \quad 1 \leq t \leq m. \end{aligned}$$

Thus, by Jensen's inequality, we have

$$\begin{aligned} \mathrm{E}\{|q_2(\hat\psi) - q_2(\psi)|^2\} &\leq c(1+\log m)^{2L(K_3+1)}\mathrm{E}\left\{\frac{1}{m}\sum_{t=1}^m f_4(\xi_i,\xi_t)\right\}^2 \\ &\leq c(1+\log m)^{2L(K_3+1)}. \end{aligned}$$

Finally, by the Cauchy-Schwarz inequality, we have

$$\Delta_2 \leq [\mathrm{E}\{|q_2(\hat{\psi}) - q_2(\psi)|^2\}]^{1/2}\mathrm{P}_\psi(\mathcal{B}^c)^{1/2} \leq c(1 + \log m)^{L(K_3+1)}m^{-b\delta/2} = o(1).$$

3. Proof of Theorem 2. Recall $d(\psi) = b(\psi) - c(\psi)$, where $b(\psi) = \mathrm{E}\{a(y, \psi)\}, c(\psi) = \mathrm{E}\{a(y, \hat{\psi})\}$. Note that $b(\psi)$ is the MSPE when ψ is the true parameter vector. Denote the right side of (12) by $\hat{d}(\psi)$, which is an approximation to $d(\psi)$. Note that, in (12), $y_{[k]}$ is y generated under ψ through ξ, introduced above Theorem 2, which does not depend on ψ and is independent of $\hat{\psi}$, the estimator of ψ based on the original data. Furthermore, $\hat{\psi}_{[k]}$ is a function of $y_{[k]}$. Thus, the summand in (12) is a function of $\xi_{[k]}$, the kth Monte-Carlo copy of ξ, and ψ. Denote the summands by $\Delta(\xi_{[k]}, \psi), 1 \leq k \leq K$. Then, we have $d(\psi) = \mathrm{E}_{\mathrm{d}}\{a(y, \psi) - a(y, \hat{\psi})\} = \mathrm{E}_{\mathrm{mc}}\{\Delta(\xi, \psi)\}$, where E_{d} denotes expectation with respect to the data, and E_{mc} that with respect to the Monte-Carlo simulation, that is, with respect to ξ (see the paragraph above Theorem 2). The two expectations are equal because y can be generated the same way as $y_{[k]}$, through ξ and given the same ψ. It follows that

$$\mathrm{E}_{\mathrm{mc}}\{\hat{d}(\psi)\} = \frac{1}{K}\sum_{k=1}^{K}\mathrm{E}_{\mathrm{mc}}\{\Delta(\xi_{[k]}, \psi)\} = \mathrm{E}_{\mathrm{mc}}\{\Delta(\xi, \psi)\} = d(\psi). \tag{A.16}$$

The Sumca estimator, (13), can now be expressed as

$$\widehat{\mathrm{MSPE}}_K = a(y, \hat{\psi}) + \hat{d}(\hat{\psi}), \tag{A.17}$$

where, in $\hat{d}(\hat{\psi})$, that is, the summand in (13), $y_{[k]}$ is generated via $\xi_{[k]}$ and $\psi = \hat{\psi}$ as described above Theorem 2. In other words, the summands in (13) are $\Delta(\xi_{[k]}, \hat{\psi})$.

By the proof of Theorem 1, we have $\mathrm{E}_{\mathrm{d}}\{d(\hat{\psi}) - d(\psi)\} = o(m^{-1})$. Thus, we have

$$\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\psi)\} = \mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})\} + o(m^{-1}), \tag{A.18}$$

where E denotes expectation with respect to both the data and Monte-Carlo simulation. Note that $d(\hat{\psi}) - d(\psi)$ depends only on y and not ξ.

On the other hand, we have $\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})\} = \mathrm{E}[\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})|\hat{\psi}\}]$. For any given value of ψ, we have, by the independence of ξ and y, and (A.16), $\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})|\hat{\psi} = \psi\} = \mathrm{E}\{\hat{d}(\psi) - d(\psi)|\hat{\psi} = \psi\} = \mathrm{E}_{\mathrm{mc}}\{\hat{d}(\psi) - d(\psi)\} = 0$. Note that $\hat{d}(\psi) - d(\psi)$ depends only on ξ and not y; thus, $\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})|\hat{\psi}\} = 0$, hence

$$\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\hat{\psi})\} = 0. \tag{A.19}$$

Combining (A.18), (A.19), we have $\mathrm{E}\{\hat{d}(\hat{\psi}) - d(\psi)\} = o(m^{-1})$. Therefore, by (A.17), we have $\mathrm{E}(\widehat{\mathrm{MSPE}}_K) = \mathrm{E}\{a(y, \hat{\psi})\} + d(\psi) + \mathrm{E}\{\hat{d}(\hat{\psi}) - d(\psi)\} = c(\psi) + b(\psi) - c(\psi) + o(m^{-1}) = b(\psi) + o(m^{-1}) = \mathrm{MSPE} + o(m^{-1})$.

4. An example: Exponential convergence rate in model selection. Consider a simple case of model selection via hypothesis testing in the i.i.d. case. Suppose that $X_1, \dots, X_n$ are independent and distributed as $N(\mu, 1)$. There are two candidate models, $M_0 : \mu = 0$ and $M_1 : \mu \neq 0$. Model selection is done by testing the hypothesis $\mathrm{H}_0 : \mu = 0$ vs $\mathrm{H}_1 : \mu \neq 0$. Consider the t-test, which rejects H_0 if $|t| = \sqrt{n}|\bar{X}| > z_{\alpha/2}$, where $\bar{X}$ is the sample mean and $z_{\alpha/2}$ the $\alpha/2$ critical value so that $\mathrm{P}(Z > z_{\alpha/2}) = \alpha/2$, α being the level of significance and $Z \sim N(0, 1)$. If H_0 is rejected, M_1 is selected; otherwise, M_0 is selected. Thus, $\hat{M} = M_0$ if $\sqrt{n}|\bar{X}| \le z_{\alpha/2}$, and $\hat{M} = M_1$ if $\sqrt{n}|\bar{X}| > z_{\alpha/2}$.

Suppose that $M_{\mathrm{o}} = M_1$. Then, $\psi_{\mathrm{o}} = \mu$, whose parameter space is $(-\infty, \infty) \setminus \{0\}$. Define a regularized estimator of μ as follows. Let $\hat{\mu} = \bar{X}$ if $|\bar{X}| \le a_n$, $\hat{\mu} = -a_n$ if $\bar{X} < -a_n$, and $\hat{\mu} = a_n$ if $\bar{X} > a_n$, where a_n is a sequence of positive constants such that $\lim_{n\to\infty} a_n = \infty$. It is easy to see that $\hat{\mu}$ is a consistent estimator of μ, whose value belongs to the parameter space with probability one.

It can be shown that $\mathrm{P}(\hat{M} \neq M_{\mathrm{o}}) = \Phi(z_{\alpha/2} - \sqrt{n}\mu) - \Phi(-z_{\alpha/2} - \sqrt{n}\mu)$, where $\Phi(\cdot)$ is the cdf of $N(0, 1)$. It follows, by the integral mean value theorem, that

$$
\begin{aligned}
\mathrm{P}(\hat{M} \neq M_{\mathrm{o}}) &= \int_{-z_{\alpha/2}-\sqrt{n}\mu}^{z_{\alpha/2}-\sqrt{n}\mu} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\
&= \sqrt{\frac{2}{\pi}} z_{\alpha/2} \exp\left\{-\frac{(b-\sqrt{n}\mu)^2}{2}\right\} \\
&= \left(\sqrt{\frac{2}{\pi}} z_{\alpha/2} e^{-b^2/2}\right) \exp\left\{-\frac{\mu^2}{2}\left(1 - \frac{2b}{\mu\sqrt{n}}\right) n\right\} \\
&\le c \exp\left(-\frac{\mu^2}{4} n\right)
\end{aligned}
$$

for some constant $c > 0$, if n is large, where $b \in [-z_{\alpha/2}, z_{\alpha/2}]$.

5. Proof of Theorem 3. Denote the right side of (18) by $r(\hat{\psi}_{\mathrm{f}})$. Then, we have

$$
\mathrm{E}(\widehat{\mathrm{MSPE}}) = \mathrm{E}(\widehat{\mathrm{MSPE}}_1) + \mathrm{E}\{r(\hat{\psi}_{\mathrm{f}})\}. \tag{A.20}
$$

For any $\psi_{\mathrm{f}} \in \Psi_{\mathrm{f},m}$, we have, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
|r(\psi_{\mathrm{f}})| &\le 2\|h_1(y, \hat{\psi}_{\mathrm{f}}) - h_1(y, \psi_{\mathrm{f}})\|_2 \|(\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}) 1_{(\hat{M} \neq M_{\mathrm{o}})}\|_2 \\
&= 2\|h_1(y, \hat{\psi}_{\mathrm{f}}) - h_1(y, \psi_{\mathrm{f}})\|_2 \|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}\|_2, \qquad \text{(A.21)}
\end{aligned}
$$

because $\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}} = 0$ when $\hat{M} = M_{\mathrm{o}}$. The first $\|\cdot\|_2$ on the right side of (A.21) is uniformly bounded by cm^{δ} for $\psi_{\mathrm{f}} \in \Psi_{\mathrm{f},m}$ by condition (i). It follows that $|r(\hat{\psi}_{\mathrm{f}})| \le 2cm^{\delta}\|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}\|_2$ by the definition of $\hat{\psi}_{\mathrm{f}}$, hence $|\mathrm{E}\{r(\hat{\psi}_{\mathrm{f}})\}| = o(m^{-1})$ by condition (iii).

Next, by similar arguments as in the proof of Theorem 1, it can be shown that

$$\mathrm{E}(\widehat{\mathrm{MSPE}}_1) = \mathrm{E}(\hat{\theta}_{\mathrm{o}} - \theta)^2 + o(m^{-1}). \tag{A.22}$$

Finally, we have $\{(\hat{\theta}_{\mathrm{o}} - \theta)^2 - (\hat{\theta}_{\mathrm{P}} - \theta)^2\}1_{(\hat{M} \neq M_0)} = (\hat{\theta}_{\mathrm{o}} - \theta + \hat{\theta}_{\mathrm{P}} - \theta)(\hat{\theta}_{\mathrm{o}} - \hat{\theta}_{\mathrm{P}})1_{(\hat{M} \neq M_{\mathrm{o}})}$. Thus, we have, again by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left|\mathrm{E}\left[\{(\hat{\theta}_{\mathrm{o}} - \theta)^2 - (\hat{\theta}_{\mathrm{P}} - \theta)^2\}1_{(\hat{M} \neq M_0)}\right]\right| &\leq \|\hat{\theta}_{\mathrm{o}} - \theta + \hat{\theta}_{\mathrm{P}} - \theta\|_2 \|(\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}})1_{(\hat{M} \neq M_{\mathrm{o}})}\|_2 \\
&\leq (\|\hat{\theta}_{\mathrm{o}} - \theta\|_2 + \|\hat{\theta}_{\mathrm{P}} - \theta\|_2)\|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}\|_2 \\
&\leq (2\|\hat{\theta}_{\mathrm{o}} - \theta\| + \|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}\|_2))\|\hat{\theta}_{\mathrm{P}} - \hat{\theta}_{\mathrm{o}}\|_2 \\
&= \{O(m^{\delta}) + o(m^{-1-\delta})\}o(m^{-1-\delta}) \\
&= o(m^{-1}), \qquad \text{(A.23)}
\end{aligned}$$

by conditions (ii) and (iii). By (A.23), we have

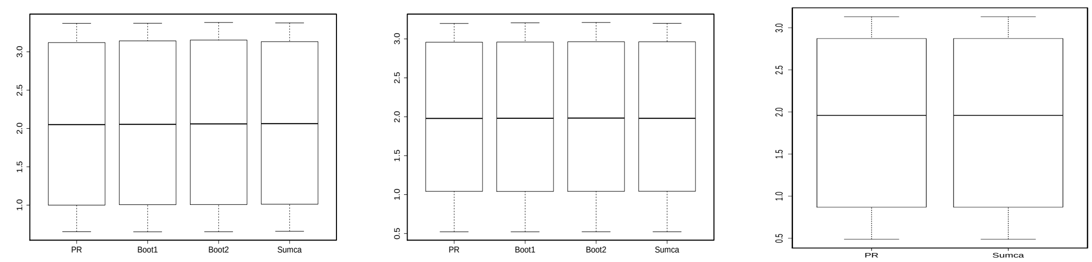
$$\begin{aligned}
\mathrm{MSPE} &= \mathrm{E}(\hat{\theta}_{\mathrm{P}} - \theta)^2 \\
&= \mathrm{E}\{\hat{\theta}_{\mathrm{P}} - \theta)^2 1_{(\hat{M} \neq M_{\mathrm{o}})}\} + \mathrm{E}\{(\hat{\theta}_{\mathrm{o}} - \theta)^2 1_{(\hat{M} = M_{\mathrm{o}})}\} \\
&= \mathrm{E}(\hat{\theta}_{\mathrm{o}} - \theta)^2 + \mathrm{E}\left[\{(\hat{\theta}_{\mathrm{P}} - \theta)^2 - (\hat{\theta}_{\mathrm{o}} - \theta)^2\}1_{(\hat{M} \neq M_{\mathrm{o}})}\right] \\
&= \mathrm{E}(\hat{\theta}_{\mathrm{o}} - \theta)^2 + o(m^{-1}) \qquad \text{(A.24)}
\end{aligned}$$

The conclusion then follows by combining (A.20), the proved fact that $\mathrm{E}\{r(\hat{\psi}_{\mathrm{f}})\} = o(m^{-1})$, (A.22), and (A.24).

A.2 Additional empirical results

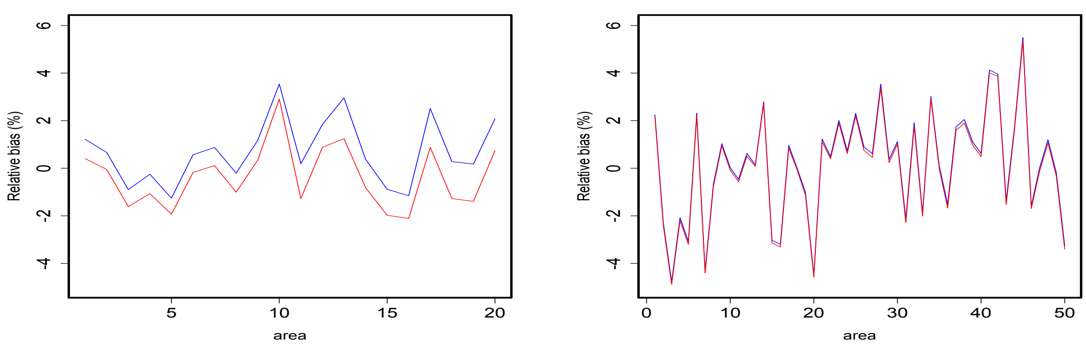
1. Further detail about Section 5.1. Figure A.1 shows performance of different MSPE estimators in terms of variability. Figure A.2 shows closeness of Sumca and PR estimates in terms of the area-specific %RB; for $m = 50$ the values are hardly distinguishable.

2. Mixed logistic model with BIC model selection. We carry out another simulation to investigate a case of PMS (see Section 3). This time, the context is a mixed logistic model (see Section 4.3), and the method for model selection is the Bayesian information criterion (BIC; Schwarz 1978). We consider the following mixed logistic model: $\mathrm{logit}(p_{ij}) = x_{ij}'\beta + v_i$, $i = 1, ..., m; j = 1, ..., n_i$, where $v_i \sim N(0, A)$ with $A = 1, m = 100, n_i = 5$, and $p_{ij} = \mathrm{P}(y_{ij} = 1|v_i)$, y_{ij} being the binary response. There are two candidate models for $x_{ij}'\beta$, Model 1: $x_{ij}'\beta = \beta_0$; and Model 2: $x_{ij}'\beta = \beta_0 + \beta_1 x_{ij}$. We consider $x_{ij} = x_i$, generated from the Uniform$[0, 1]$ distribution, and fixed throughout.



(a) (b) (c)

Figure A.1: Boxplots of MSPE estimates using Sumca, DB (Boot1 and Boot2) and PR methods: (a) $m = 20$; (b) $m = 50$; (c) $m = 200$.



(a) (b)

Figure A.2: % RB of MSPE estimates using Sumca and Prasad-Rao methods: (a) $m = 20$ and (b) $m = 50$. Blue color: Sumca estimator; Red color: Prasad-Rao estimator.

We generate the data under Model 2 (the true model) with $\beta_0 = \beta_1 = 1$. We then carry out the BIC procedure to select the model; after the model is selected, we obtain the small area predictor of $p_{ij} = p_i$ under the selected model. The Sumca estimator is computed with $K = 100$. Note that, according to Section 4, the Sumca estimator is always computed under the full model, which is Model 2 in this case. As an alternative method to Sumca, we obtain the JLW MSPE estimator under the selected model (in a way, this is similar to the DHM method, discussed in Sections 3.2 and 4.2). We then calculate the EMSPE of $\hat{p}_i^{(r)}$ over the simulated data sets r by (31). The % RB of MSPE estimation, (32), is used to evaluate the performance of the two PMS MSPE estimators, Sumca and JLW under the selected model. The boxplots of %RB are presented in Figure A.3.

It appears that the Sumca estimator performs slightly better than JLW under the selected model in terms of %RB. Note that this is a situation in favor of JLW. In fact, according to our simulation results (see Jiang and Torabi 2019), JLW performs slightly better, overall, than Sumca when model selection is not involved. The current simulated example shows that, in a PMS situation, it is the other way around, that is, Sumca performs slightly better than JLW under the selected model.

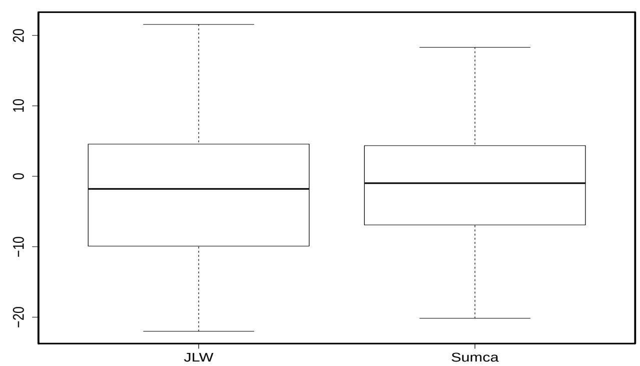


Figure A.3: Boxplots of % RB: Sumca vs JLW under the selected model

3. Application: Minority health insurance data. We evaluate the performance of Sumca and JLW MSPE estimators using a real data set that utilizes a mixed logistic model. Additional real-data applications are provided in Jiang and Torabi (2019).

Ghosh *et al.* (2009) considered small domain estimation of the proportion of persons without health insurance for different minority groups in the Asian population. The small domains were constructed on the basis of age, sex, race, and region where the persons belonged. The authors used data provided by National Health Interview Survey (NHIS) for the year 2000, which report the individual level binary responses on whether a person has health insurance, along with his or her individual level covariates. The Asian group is composed of four categories: Chinese, Filipino, Asian Indian, and Others such as Korean, Vietnamese, Japanese, Hawaiian, Samoan, Guamanian, etc. These individuals were assigned to specific domains depending on their age, gender, race, and region they came from. There were three age-groups (0-17, 18-64, 65+), two groups for gender, four regions depending on the size of the Metropolitan Statistical Area ($< 499,999; 500,000 - 999,999; 1,000,000 - 2,499,999; > 2,500,000$), and four groups for race. The total number of domains is then $96\ (= 3 \times 2 \times 4 \times 4)$. The sample sizes for some domains of a targeted minority Asian population were not large enough to produce reliable estimates; in fact, only 19 out of the 96 domains have non-zero sample sizes. In order to address this issue, Ghosh *et al.* (2009) employed both Hierarchical Bayes and empirical Bayes methodologies to obtain small domain estimates and also to find the associated measures of precision. In particular, they considered the following model: $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + v_i$, $i = 1, ..., 96; j = 1, ..., n_i$, where $v_i \overset{\text{i.i.d.}}{\sim} N(0, A), p_{ij} = \text{P}(y_{ij} = 1|v_i)$ with $y_{ij} = 1$ or 0 depending on whether or not the jth individual in the ith small domain does not have health insurance; $x_{ij1}, x_{ij2}, x_{ij3}$ are family size, educational level, and total family income of the jth unit in the ith small domain, respectively. We estimate the model parameters as $\hat{\beta}_0 = -4.19, \hat{\beta}_1 = 0.62, \hat{\beta}_2 = 0.005, \hat{\beta}_3 = 0.12, \hat{A} = 0.004$. We then compute the EBP $\hat{p}_i$, where $p_i = h(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + v_i)$ with $h(u) = e^u/(1 + e^u)$ and (x_{i1}, x_{i2}, x_{i3}) being the average of $(x_{ij1}, x_{ij2}, x_{ij3})$ over domain i, and the Sumca estimate $\widehat{\text{MSPE}}_{i,K}$ using (13) with $K = 100$ and JLW MSPE estimate $\widehat{\text{MSPE}}_{i,\text{JLW}}$, as well as the bootstrap MSPE estimator $\widehat{\text{MSPE}}_{i,\text{Boot}}$ (see Section 4.3). Figure A.4 shows boxplots of the square roots of the MSPE estimates. It is observed that the Sumca estimates are relatively (much) smaller than the JLW and bootstrap estimates. This seems to be consistent with our simulation results (see Jiang and Torabi 2019) that JLW and bootstrap tend to over-estimate the MSPE. It is known that the bootstrap MSPE estimator is only first-order unbiased (e.g., Hall and Maiti 2006b), so the overestimation by $\widehat{\text{MSPE}}_{i,\text{Boot}}$ is not surprising. As for JLW, although it is known to be second-order unbiased (Jiang *et al.* 2002), its application in this case involves Monte-Carlo simulation [to compute (30) in Section 4.3]. It is possible that the Monte-Carlo sample size, $K = 100$, is not large enough to ensure that the Monte-Carlo approximations to terms like (30) have combined error of the order $o(m^{-1})$. Note that $K = 100$ is chosen partially due to the computational intensity of JLW, which involves evaluation of many $O(m)$ terms [see (29)], and partially due to our intention to make a fair comparison, as $K = 100$ is also used for computing the Sumca estimator.

Additional References

Cox, D. R. and Hinkley, D. V. (1973), *Theoretical Statistics*, Chapman & Hall, London.

Das, K. and Jiang, J. and Rao, J. N. K. (2004), Mean squared error of empirical predictor, *Ann. Statist.* 32, 818-840.

Ghosh, M., Kim, D., Sinha, K., Maiti, T., Katzoff, M. and Parsons, V.L. (2009), Hierarchical and empirical Bayes small domain estimation of the proportion of persons without

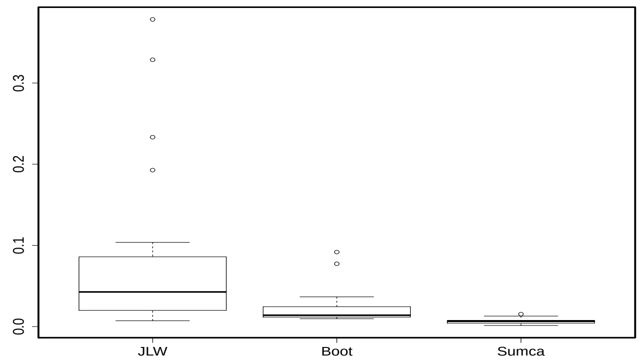


Figure A.4: Boxplots of square roots of MSPE estimates for Health Insurance data

health insurance for minority subpopulations, *Surv. Methodol.* 35, 53-66.

Pfanzagl, J. (1980), Asymptotic expansions in parametric statistical theory, in *Developments in Statistics* 3, 1-97.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Statist.* 6, 461-464.

Shao, J. (2003), *Mathematical Statistics*, 2nd ed., Springer, New York.