# Supplementary Material on
## *"An Averaging Estimator for Two Step M Estimation in Semiparametric Models"*

Ruoyao Shi[1]

September, 2022

### Appendix B Proofs of the Lemmas in Appendix A

#### Proof of Lemma A.1

*Proof.* The following proves inequality (A.8). By the definition of supremum and the definition of $Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP})$ in (A.6), there exists a sequence of DGPs, denoted by $\{F_n\}_{n\in\mathbb{N}}$, such that

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) = \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})].$$

The real sequence $\{\mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})]\}_{n\in\mathbb{N}}$ itself may not be convergent, but by the definition of limsup, there exists a subsequence of $\{n\}_{n\in\mathbb{N}}$, denoted by $\{p_n\}_{n\in\mathbb{N}}$, such that the corresponding real subsequence $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n}}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n}})]\}_{n\in\mathbb{N}}$ is convergent, where $\{F_{p_n}\}_{n\in\mathbb{N}}$ denotes the subsequence of DGPs corresponding to $\{p_n\}_{n\in\mathbb{N}}$, then

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) = \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n}}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n}})]. \tag{B.1}$$

Now consider the sequence of $k$-dimensional vectors $\{p_n^{1/2}\delta_{F_{p_n}}\}_{n\in\mathbb{N}}$, and let $\{p_n^{1/2}\delta_{F_{p_n},\iota}\}_{n\in\mathbb{N}}$ ($\iota = 1, \ldots, k$) denote their $\iota$th coordinates. For $\iota = 1$, one has either (i) $\limsup_{n\to\infty} \left|p_n^{1/2}\delta_{F_{p_n},\iota}\right| < \infty$, or (ii) $\limsup_{n\to\infty} \left|p_n^{1/2}\delta_{F_{p_n},\iota}\right| = \infty$. For case (i), there exists some subsequence $\{p_{n,\iota}\}_{n\in\mathbb{N}}$ such that $p_{n,\iota}^{1/2}\delta_{F_{p_{n,\iota}},\iota} \to d_\iota$ for some $d_\iota \in \mathbb{R}$, by the definition of limsup. For case (ii), there exists some subsequence $\{p_{n,\iota}\}_{n\in\mathbb{N}}$ such that $p_{n,\iota}^{1/2}\delta_{F_{p_{n,\iota}},\iota} \to \infty$ or $-\infty$, by the definition of limsup. In both cases, therefore, there exists some subsequence $\{p_{n,\iota}\}_{n\in\mathbb{N}}$ such that $p_{n,\iota}^{1/2}\delta_{F_{p_{n,\iota}},\iota} \to d_\iota$ for some $d_\iota \in \mathbb{R}_\infty$. Since $k$ is finite, one can sequentially apply the same argument to all coordinates $\iota = 2, \ldots, k$ and let the resulting subsequence be denoted by $\{p_{n,k}\}_{n\in\mathbb{N}}$, which satisfies $p_{n,k}^{1/2}\delta_{F_{p_{n,k}}} \to d$ for some $d \in \mathbb{R}_\infty^k$. Next consider $\{S(F_{p_{n,k}})\}_{n\in\mathbb{N}}$, the sequence of nuisance parameter vectors in $\mathcal{S}$ induced by the DGPs $\{F_{p_{n,k}}\}_{n\in\mathbb{N}}$. $\{S(F_{p_{n,k}})\}_{n\in\mathbb{N}}$ itself may not be convergent, but since $\mathcal{S}$ is compact by Condition 3(i), then there exists a convergent subsequence, denoted by $\{S(F_{p_n^*})\}_{n\in\mathbb{N}}$, such that $S(F_{p_n^*}) \longrightarrow s^*$ with some $s^* \in \mathcal{S}$. Moreover, by Condition 3(ii), there exists a DGP $F^*$ in $\mathcal{F}$ such that $S(F^*) = s^*$. This shows that there exists some subsequence $\{p_n^*\}_{n\in\mathbb{N}}$ of $\{p_n\}_{n\in\mathbb{N}}$ such that

$$p_n^{*1/2}\delta_{F_{p_n^*}} \to d \text{ for some } d \in \mathbb{R}_\infty \text{ and } S(F_{p_n^*}) \longrightarrow S(F^*) \text{ for some } F^* \in \mathcal{F}. \tag{B.2}$$

Note that for any subsequence of $\{p_n\}_{n\in\mathbb{N}}$, the limit of the right hand side in (B.1) remains the same, which implies

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) = \lim_{n\to\infty} \mathbb{E}_{F_{p_n^*}}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n^*}}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n^*}})]. \tag{B.3}$$

The definitions of $\ell(\hat{\beta}_n, \beta)$ in (2.1) and of $\ell_\zeta(\hat{\beta}_n, \beta)$ in (3.10), as well as (A.3) suggest that in order to prove (A.8), one needs to link the right hand side of (B.3) with $R_\zeta(u_{F,d})$ and $\bar{R}_\zeta(u_{F,d})$ defined in (A.1) and

[1] Ruoyao Shi: Department of Economics, UC Riverside, USA. Email: ruoyao.shi@ucr.edu.

(A.2). First consider the case where $\|d\| < \infty$ in (B.2). By Condition 2(i) and Lemma 1(i),

$$p_n^{*1/2}(\hat{\beta}_{n,SP} - \beta_{F_{p_n^*}}) \xrightarrow{d.} \xi_{F,SP} \text{ and } p_n^{*1/2}(\hat{\beta}_{n,\hat{w}_n} - \beta_{F_{p_n^*}}) \xrightarrow{d.} \bar{\xi}_{F,d},$$

which combined with the continuous mapping theorem implies that

$$\ell(\hat{\beta}_{n,SP}, \beta_{F_{p_n^*}}) \xrightarrow{d.} \xi_{F,SP}'\Upsilon\xi_{F,SP} \text{ and } \ell(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n^*}}) \xrightarrow{d.} \bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d}.$$

Since $\Upsilon$ is positive semi-definite, $\xi_{F,SP}'\Upsilon\xi_{F,SP}$ and $\bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d}$ are both nonnegative. Note that the function $f(x) \equiv \min\{x, \zeta\}$ is a bounded continuous function of $x \geq 0$ for fixed positive $\zeta$. Applying the Portmanteau lemma (e.g., Lemma 2.2 in Van der Vaart, 2000) and invoking (A.1) and (A.2), one gets

$$\mathbb{E}_{F_{p_n^*}}\left[\ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n^*}})\right] \to R_\zeta(u_{F^*,d}) \text{ and } \mathbb{E}_{F_{p_n^*}}\left[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n^*}})\right] \to \bar{R}_\zeta(u_{F^*,d}). \tag{B.4}$$

Next consider the case where $\|d\| = \infty$ in (B.2). By Condition 2(ii) and Lemma 1(ii),

$$p_n^{*1/2}(\hat{\beta}_{n,SP} - \beta_{F_{p_n^*}}) \xrightarrow{d.} \xi_{F,SP} \text{ and } p_n^{*1/2}(\hat{\beta}_{n,\hat{w}_n} - \beta_{F_{p_n^*}}) \xrightarrow{d.} \xi_{F,SP}.$$

Using the same argument, one also gets (B.4) in this case. Combining (A.3), (B.3) and (B.4), one can unify the two cases and write

$$\begin{aligned} Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) &= r_\zeta(u_{F^*,d}), \text{ for some } F^* \in \mathcal{F} \text{ and some } d \in \mathbb{R}_\infty^k \\ &\leq \max\left\{\sup_{u_{F,d}\in\mathcal{U}} r_\zeta(u_{F,d}), \sup_{u_{F,d}\in\mathcal{U}_\infty} r_\zeta(u_{F,d})\right\} \\ &= \max\left\{\sup_{u_{F,d}\in\mathcal{U}} r_\zeta(u_{F,d}), 0\right\}. \end{aligned}$$

This proves (A.8).

The proof of (A.9) follows the same argument and hence is omitted here. ■

**Proof of Lemma A.2**

*Proof.* The following proves inequality (A.10). By the definition of $\mathcal{U}$ in (3.25), $\|d\| < \infty$ and $\delta_F = 0$ for any $F \in \mathcal{F}$ such that $u_{F,d} \in \mathcal{U}$. For any $u_{F,d} \in \mathcal{U}$, let $N_{\epsilon_F}$ denote the smallest $n$ such that $n^{-1/2}\|d\| < \epsilon_F$, where $\epsilon_F$ satisfies Condition 3(ii). Then by Condition 3(ii), for each $n \geq N_{\epsilon_F}$, there is an $F_n \in \mathcal{F}$ with $\delta_{F_n} = n^{-1/2}d$ and $\|\bar{S}(F_n) - \bar{S}(F)\| \leq n^{-\kappa/2}C\|d\|^\kappa$ for some $C, \kappa > 0$. For each $n < N_{\epsilon_F}$, let $F_n = F$. Thus, a sequence of DGPs $\{F_n\}_{n\in\mathbb{N}}$ in $\mathcal{F}$ satisfying $n^{1/2}\delta_{F_n} \to d$ and $\bar{S}(F_n) \to \bar{S}(F)$ is constructed for any $u_{F,d} \in \mathcal{U}$. Recalling the definition of $\bar{S}(F)$ in (3.17), this immediately implies that for such $\{F_n\}_{n\in\mathbb{N}}$,

$$n^{1/2}\delta_{F_n} \to d \in \mathbb{R}^k,\ V_{F_n,SP} \to V_{F,SP},\ C_{F_n} \to C_F, \text{ and } V_{F_n,P} \to V_{F,P}. \tag{B.5}$$

The real sequence $\{\mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})]\}_{n\in\mathbb{N}}$ that corresponds to $\{F_n\}_{n\in\mathbb{N}}$ may not be convergent, but by the definition of lim sup, there exists a subsequence $\{p_n\}_{n\in\mathbb{N}}$ of $\{n\}_{n\in\mathbb{N}}$ such that the corresponding real sequence $\{\mathbb{E}_{F_{p_n}}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n}}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n}})]\}_{n\in\mathbb{N}}$ is convergent and

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n}}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n}})] = \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})]. \tag{B.6}$$

Since $\{p_n\}_{n\in\mathbb{N}}$ is a subsequence of $\{n\}_{n\in\mathbb{N}}$, (B.5) implies that

$$n^{1/2}\delta_{F_{p_n}} \to d \in \mathbb{R}^k,\ V_{F_{p_n},SP} \to V_{F,SP},\ C_{F_n} \to C_F, \text{ and } V_{F_{p_n},P} \to V_{F,P}. \tag{B.7}$$

Combined with Condition 2(i) and Lemma 1(i), this implies that

$$p_n^{1/2}(\hat{\beta}_{n,SP} - \beta_{F_{p_n}}) \xrightarrow{d.} \xi_{F,SP}, \text{ and } p_n^{1/2}(\hat{\beta}_{n,\hat{w}_n} - \beta_{F_{p_n}}) \xrightarrow{d.} \bar{\xi}_{F,d},$$

which, combined with the continuous mapping theorem, in turn implies that

$$\lim_{n\to\infty} \mathbb{E}_{F_{p_n}}\left[\ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_{p_n}})\right] = R_\zeta(u_{F,d}), \text{ and } \lim_{n\to\infty} \mathbb{E}_{F_{p_n}}\left[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_{p_n}})\right] = \bar{R}_\zeta(u_{F,d}). \tag{B.8}$$

This, combined with (B.6), the definition of $Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP})$ in (A.6), the definition of supremum and the definition of $r(u_{F,d})$ in (A.3), implies that for any $u_{F,d} \in \mathcal{U}$,

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})] = r(u_{F,d}),$$

which further implies that

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) \geq \sup_{u_{F,d}\in\mathcal{U}} r(u_{F,d}). \tag{B.9}$$

On the other hand, by the definition of $\mathcal{U}_\infty$ in (3.26), for any $u_{F,d} \in \mathcal{U}_\infty$, $\|d\| = \infty$ and either (i) $\delta_F = 0$ or (ii) $\|\delta_F\| > 0$. For case (i), let $\ell_k$ be a $k \times 1$ vector of ones and let $N_{\epsilon_F}$ denote the smallest $n$ such that $n^{-1/4}\|\ell_k\|^{1/2} = n^{-1/4}k^{1/2} < \epsilon_F$, where $\epsilon_F$ satisfies Condition 3(ii). Then by Condition 3(ii), for each $n \geq N_{\epsilon_F}$, there is an $F_n \in \mathcal{F}$ with $\delta_{F_n} = n^{-1/4}\ell_k$ and $\|\bar{S}(F_n) - \bar{S}(F)\| \leq Cn^{-\kappa/4}k^{\kappa/2}$ for some $C, \kappa > 0$. For each $n < N_{\epsilon_F}$, let $F_n = F$. For case (ii), let $F_n = F$ for all $n$. Thus, a sequence of DGPs $\{F_n\}_{n\in\mathbb{N}}$ in $\mathcal{F}$ satisfying $n^{1/2}\delta_{F_n} \to \infty$, $\delta_{F_n} \to F$ and $\bar{S}(F_n) \to \bar{S}(F)$ is constructed for any $u_{F,d} \in \mathcal{U}_\infty$, regardless of whether $\delta_F = 0$ or $\|\delta_F\| > 0$. Recalling the definition of $\bar{S}(F)$ in (3.17), this immediately implies that for such $\{F_n\}_{n\in\mathbb{N}}$,

$$\|n^{1/2}\delta_{F_n}\| \to \infty,\ V_{F_n,SP} \to V_{F,SP},\ C_{F_n} \to C_F, \text{ and } V_{F_n,P} \to V_{F,P}.$$

Then a similar argument used to show (B.6) - (B.8) can be applied to show that there exists a subsequence $\{p_n\}_{n\in\mathbb{N}}$ of $\{n\}_{n\in\mathbb{N}}$ such that (B.6) and (B.8) are satisfied, with the help of Condition 2(ii) and Lemma 1(ii). Combining this with the definition of $Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP})$ in (A.6), the definition of supremum and the definition of $r(u_{F,d})$ in (A.3), one gets that for any $u_{F,d} \in \mathcal{U}_\infty$,

$$Asy\overline{RD}_\zeta(\hat{\beta}_{n,\hat{w}_n}, \hat{\beta}_{n,SP}) \geq \limsup_{n\to\infty} \mathbb{E}_{F_n}[\ell_\zeta(\hat{\beta}_{n,\hat{w}_n}, \beta_{F_n}) - \ell_\zeta(\hat{\beta}_{n,SP}, \beta_{F_n})] = 0. \tag{B.10}$$

(A.10) immediately follows inequalities (B.9) and (B.10).

The proof of (A.11) follows the same argument and hence is omitted here. ■

**Proof of Lemma A.3**

*Proof.* For any $F \in \mathcal{F}$, since $\xi_{F,SP} \sim \mathcal{N}(0_{k\times 1}, V_{F,SP})$ by Condition 2, one gets

$$\xi'_{F,SP}\Upsilon\xi_{F,SP} \overset{d.}{=} \mathcal{Z}'V_{F,SP}^{1/2}\Upsilon V_{F,SP}^{1/2}\mathcal{Z},$$

where $\mathcal{Z} \sim \mathcal{N}(0_{k\times 1}, I_{k\times k})$. By Condition 3(i), and because $\Upsilon$ is a fixed matrix, there exists some constant $C$ such that

$$\sup_{F\in\mathcal{F}} \rho_{\max}\left(V_{F,SP}^{1/2}\Upsilon V_{F,SP}^{1/2}\right) \leq C.$$

This implies that

$$\sup_{u_{F,d}\in\mathcal{U}} \mathbb{E}\left[\left(\xi_{F,SP}'\Upsilon\xi_{F,SP}\right)^2\right] \leq \sup_{u_{F,d}\in\mathcal{U}} \rho_{\max}^2\left(V_{F,SP}^{1/2}\Upsilon V_{F,SP}^{1/2}\right)\cdot\mathbb{E}[(\mathcal{Z}'\mathcal{Z})^2] \leq C,$$

where the second inequality holds because $\mathcal{Z} \sim \mathcal{N}(0_{k\times 1}, I_{k\times k})$ and that $V_{F,SP}$ does not depend on $d$. This proves (A.12).

By the definition of $\bar{\xi}_{F,d}$ in (3.22) and that of $\tilde{\xi}_F$ in Condition 2(i), Cauchy-Schwarz inequality and the simple inequality $2|ab| \leq a^2+b^2$ for any real numbers $a$ and $b$, one gets

$$\begin{aligned}\bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d} &\leq 2\xi_{F,SP}'\Upsilon\xi_{F,SP} + 2w_F^2\left(\xi_{F,P}+d-\xi_{F,SP}\right)'\Upsilon\left(\xi_{F,P}+d-\xi_{F,SP}\right)\\ &= 2\xi_{F,SP}'\Upsilon\xi_{F,SP} + 2w_F^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right),\end{aligned}\tag{B.11}$$

where $D$ and $\tilde{d}$ are defined in (A.21). Combining (B.11) and the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers $a$ and $b$, one gets

$$\begin{aligned}\left(\bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d}\right)^2 \leq& 8\left(\xi_{F,SP}'\Upsilon\xi_{F,SP}\right)^2 + 8\left[w_F^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right)\right]^2\\ \leq& C + 8\left[w_F^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right)\right]^2,\end{aligned}\tag{B.12}$$

where the second inequality is by (A.12). By the definitions of $w_F$ in (3.21) and that of $A_F$ and $B_F$ in (3.20), one has

$$w_F^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right) = \frac{[\mathrm{tr}(A_F)]^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right)}{\left[\mathrm{tr}(B_F)+\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right)\right]^2} \leq C\mathrm{tr}(A_F) = C\mathrm{tr}(\Upsilon V_{F,SP}) - C\mathrm{tr}(\Upsilon C_F),$$

where the inequality follows by $\mathrm{tr}(A_F) > 0$, $\mathrm{tr}(B_F) > 0$ and that $\Upsilon$ being positive semi-definite implies $(\tilde{\xi}_F+\tilde{d})'D(\tilde{\xi}_F+\tilde{d}) \geq 0$. Combined with the simple inequality $(a+b)^2 \leq 2(a^2+b^2)$, this implies that

$$\begin{aligned}\mathbb{E}\left[w_F^2\left(\tilde{\xi}_F+\tilde{d}\right)' D\left(\tilde{\xi}_F+\tilde{d}\right)\right]^2 &\leq 2C[\mathrm{tr}(\Upsilon V_{F,SP})]^2 + 2C[\mathrm{tr}(\Upsilon C_F)]^2\\ &\leq 2C[\mathrm{tr}(\Upsilon V_{F,SP})]^2 + 2C[\mathrm{tr}(\Upsilon V_{F,SP})]^2 \leq C,\end{aligned}\tag{B.13}$$

where the second inequality holds by Condition 2(i), Condition 3(i) and that Cauchy-Schwarz inequality implies $C_F \leq \max\{V_{F,SP}, V_{F,P}\}$ for any $F \in \mathcal{F}$. Together, (B.12) and (B.13) imply (A.13), since the upper bound does not depend on $F$. ■

**Proof of Lemma A.4**

*Proof.* First note that

$$\sup_{u_{F,d}\in\mathcal{U}} \left|\mathbb{E}\left[\min\{\bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d}, \zeta\} - \bar{\xi}_{F,d}'\Upsilon\bar{\xi}_{F,d}\right]\right|$$

$$
\begin{aligned}
&= \sup_{u_{F,d}\in\mathcal{U}} \left|\mathbb{E}\left[\left(\zeta - \bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d}\right)\mathbb{I}\left\{\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d} > \zeta\right\}\right]\right| \\
&\leq \sup_{u_{F,d}\in\mathcal{U}} \mathbb{E}\left[\left|\zeta - \bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d}\right| \cdot \mathbb{I}\left\{\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d} > \zeta\right\}\right] \\
&\leq \zeta \sup_{u_{F,d}\in\mathcal{U}} \mathbb{E}\left[\mathbb{I}\left\{\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d} > \zeta\right\}\right] + \sup_{u_{F,d}\in\mathcal{U}} \mathbb{E}\left[\left(\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d}\right) \cdot \mathbb{I}\left\{\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d} > \zeta\right\}\right] \\
&\leq 2\zeta^{-1} \sup_{u_{F,d}\in\mathcal{U}} \mathbb{E}\left[\left(\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d}\right)^2\right] \leq 2C\zeta^{-1}, \qquad \text{(B.14)}
\end{aligned}
$$

where the first equality is by the fact that $\min\{x,\zeta\} - x = (\zeta - x)\cdot\mathbb{I}\{x > \zeta\}$; the first inequality is by Jensen's inequality and the fact that an indicator function is always non-negative; the second inequality holds because $\zeta > 0$, $\bar{\xi}'_{F,d}\Upsilon\bar{\xi}_{F,d} \geq 0$, and the simple inequality $|a-b| \leq a+b$ for any non-negative real numbers $a$ and $b$; the third inequality holds by Markov's inequality;[2] the fourth inequality is by (A.13) in Lemma A.3.

By (A.12) in Lemma A.3 and the same argument, one can show that

$$
\sup_{u_{F,d}\in\mathcal{U}} \left|\mathbb{E}\left[\min\{\xi_{F,SP}\Upsilon\xi_{F,SP}, \zeta\} - \xi_{F,SP}\Upsilon\xi_{F,SP}\right]\right| \leq 2C\zeta^{-1}. \qquad \text{(B.15)}
$$

Combining inequalities (B.14) and (B.15), the definitions of $r_\zeta(u_{F,d})$ and $r(u_{F,d})$ in (A.4) and (A.5), and the triangular inequality, one gets $\sup_{u_{F,d}\in\mathcal{U}} |r_\zeta(u_{F,d}) - r(u_{F,d})| \leq 4C\zeta^{-1}$, which immediately implies (A.14). ■

## Appendix C Details on Section 2

**The reasons for $\hat{w}_n \in [0,1]$ with probability one.** Note that $\widehat{V}_{n,SP} + \widehat{V}_{n,P} - \widehat{C}_n - \widehat{C}'_n$ is the sample asymptotic variance-covariance matrix of $\hat{\beta}_{n,P} - \hat{\beta}_{n,SP}$ and recall that $\Upsilon$ is symmetric positive semi-definite, so the first term in the denominator of (2.2) is positive with probability one; the second term is a quadratic form with positive semi-definite $\Upsilon$, so the denominator of (2.2) is positive with probability one. Moreover, if the parametric restrictions are correctly specified or mildly misspecified, then $V_{F,SP} \geq V_{F,P}$ (Condition 2) implies $\widehat{V}_{n,SP} \geq \widehat{V}_{n,P}$ with probability one, which, together with Cauchy-Schwarz inequality, further implies $\widehat{V}_{n,SP} \geq \widehat{C}_n$.[3] Furthermore, if the parametric restrictions are severely misspecified, then $\widehat{V}_{n,SP}$, $\widehat{V}_{n,P}$ and $\widehat{C}_n$ having finite probability limits (Condition 2) implies that the second term in the denominator approaches the infinity while the other terms are finite. Together, these imply that the averaging weight $\hat{w}_n \in [0,1]$ with probability one.

**Computing asymptotic variance-covariance matrices via robust influence functions.** Let $\psi_{F,SP}(z)$ denote the non-centered influence function of $\hat{\beta}_{n,SP}$, let $\psi_{F,P}(z)$ denote that of $\hat{\beta}_{n,P}$, and let $\psi_{n,SP}(z)$ and $\psi_{n,P}(z)$ denote their sample analogs, respectively. Then

$$
\widehat{V}_{n,SP} = \frac{1}{n}\sum_{i=1}^n \psi_{n,SP}(Z_i)\psi'_{n,SP}(Z_i) - \left[\frac{1}{n}\sum_{i=1}^n \psi_{n,SP}(Z_i)\right]\cdot\left[\frac{1}{n}\sum_{i=1}^n \psi_{n,SP}(Z_i)\right]', \qquad \text{(C.1)}
$$

$$
\widehat{V}_{n,P} = \frac{1}{n}\sum_{i=1}^n \psi_{n,P}(Z_i)\psi'_{n,P}(Z_i) - \left[\frac{1}{n}\sum_{i=1}^n \psi_{n,P}(Z_i)\right]\cdot\left[\frac{1}{n}\sum_{i=1}^n \psi_{n,P}(Z_i)\right]', \qquad \text{(C.2)}
$$

[2] The first term is bounded using Chebyshev's inequality. Using the same argument as Markov's inequality, one can show that for non-negative random variable $X$ and constant $a > 0$, $\mathbb{E}[X\cdot\mathbb{I}\{X > a\}] \leq \mathbb{E}(X^2)/a$, since $\mathbb{E}(X^2) = \mathbb{E}[X^2\cdot\mathbb{I}\{X > a\}] + \mathbb{E}[X^2\cdot\mathbb{I}\{X \leq a\}] \geq \mathbb{E}[X^2\cdot\mathbb{I}\{X > a\}] \geq a\mathbb{E}[X\cdot\mathbb{I}\{X > a\}]$. Applying this result to the second term gives the desired inequality.

[3] Condition 2(i) postulates $V_{F,SP} \geq V_{F,P}$, which is the case where the averaging is meaningful, otherwise $\hat{\beta}_{n,P}$ dominates $\hat{\beta}_{n,SP}$. Allowing for $V_{F,SP} < V_{F,P}$ is also easy and discussed in Remark 2.

$$\widehat{C}_n = \frac{1}{n}\sum_{i=1}^{n}\psi_{n,SP}(Z_i)\psi'_{n,P}(Z_i) - \left[\frac{1}{n}\sum_{i=1}^{n}\psi_{n,SP}(Z_i)\right]\cdot\left[\frac{1}{n}\sum_{i=1}^{n}\psi_{n,P}(Z_i)\right]'. \tag{C.3}$$

The asymptotic variance-covariance matrix estimates in (C.1) - (C.3) can then be plugged into (2.2) to compute the averaging weight.

It is worth emphasizing that the influence functions need to be valid under *potential misspecification* (e.g., Ichimura and Lee, 2010), such that the estimators $\widehat{V}_{n,SP}$, $\widehat{V}_{n,P}$ and $\widehat{C}_n$ are consistent *regardless* of whether the parametric restrictions hold or not. In other words, they must be robust against misspecification of the parametric restrictions; otherwise the resulting averaging estimator might not conform to the asymptotic theory in Section 3. Appendix D will illustrate this point using the partially linear model in Section 4.

## Appendix D Details on Section 4

**Robust influence function in partially linear models.** In this example, the estimators are based on the objective function $Q(z,\beta,h) \equiv \frac{1}{2}[y - h_1(x_2) - (x_1 - h_2(x_2))'\beta]^2$ where $z$ represents the vector of all observed variables, so that $Q_F(\beta,h)$ in (3.1) equals to $\mathbb{E}_F[Q(Z,\beta,h)]$, where the expectation is taken with regard to the distribution $F$ of the data $Z$. Under DGP $F$, let $h_{1F}(s) \equiv \mathbb{E}_F(Y|s(X_1,X_2) = s)$ and $h_{2F} \equiv \mathbb{E}_F(X_1|s(X_1,X_2) = s)$ denote the conditional mean functions of $Y_i$ and $X_{1i}$ given $s(X_{1i},X_{2i}) = s$.[4] Since these functions do not depend on $\beta$, the general formula of the influence function of an estimator $\hat{\beta}_n$ robust to potential misspecification of $h$ can be derived using Theorem 3.3 of Ichimura and Lee (2010). Borrowing their notation, one gets

$$\begin{aligned}
\Delta_1(z) \equiv D_\beta Q(z,\beta,h) &= -[y - h_1(x_2) - (x_1 - h_2(x_2))'\beta](x_1 - h_2(x_2)),\\
D_{\beta\beta'}Q(z,\beta,h) &= (x_1 - h_2(x_2))(x_1 - h_2(x_2))',\\
V_0 &= \frac{d^2Q(\beta,h)}{d\beta d\beta'} = D_{\beta\beta'}Q(\beta,h) = \mathbb{E}[(X_1 - h_2(X_2))(X_1 - h_2(X_2))'],\\
D_hQ(z,\beta,h_F)[h] &= -[y - h_{1F}(x_2) - (x_1 - h_{2F}(x_2))'\beta](h_1(x_2) - h_2(x_2)'\beta),\\
\Gamma_1(z) &= \frac{d}{d\beta'}D_hQ(\beta,h_F)[h]\\
&= D_{\beta h}Q(\beta,h_F)[h]\\
&= \mathbb{E}_F[(X_1 - h_{2F}(X_2))'\beta(h_1(X_2) - h_2(X_2)'\beta]\\
&\quad + \mathbb{E}_F[(Y - h_{1F}(X_2) - (X_1 - h_{2F}(X_2))'\beta)h_2(X_2)]\\
&= 0,
\end{aligned}$$

where the last equality holds by the law of iterated expectations (i.e., first conditional on $X_2$). By Theorem 3.3 of Ichimura and Lee (2010), the influence function of an estimator $\hat{\beta}_n$ is

$$\begin{aligned}
\psi(z) &= -V_0^{-1}\Delta_1(z)\\
&= -\{\mathbb{E}[(X_1 - h_2(s(X_1,X_2)))\cdot(X_1 - h_2(s(X_1,X_2)))']\}^{-1}\\
&\qquad \cdot[y - h_1(s(x_1,x_2)) - (x_1 - h_2(s(x_1,x_2)))'\beta]\cdot(x_1 - h_2(x_1,x_2)).
\end{aligned} \tag{D.1}$$

[4] Here $s(X_1,X_2) = s$ is a shorthand notation to indicate conditioning on all the additively separable components of $s(X_1,X_2)$. For example, in the model (4.2) in Section 4, $s(X_1,X_2) = s$ is a shorthand for the entire vector $(X_2', X_{11}X_{21}, X_{12}X_{22}, X_{13}X_{23}, X_{14}X_{24})'$ being fixed.

Note that $\Gamma_1(z)$, the term in Ichimura and Lee (2010) that captures the impact of first step estimation error of $h$ on the asymptotic distribution of $\hat{\beta}_n$, is zero.

For the parametric estimator, $s$ is restricted to be a linear function of $x_2$ only, i.e., $s(x_1, x_2) = \theta_0 + x_2'\theta_1$, then $\hat{\beta}_{n,P}$ is just the least squares coefficient of $X_1$ in a linear regression of $Y$ on $X_1$, $X_2$, and an intercept. Under this modeling restriction, both $h_1$ and $h_2$ are also linear functions of $x_2$ only. Let $X_2^* \equiv (1, X_2')'$, then standard results of linear regressions imply that $h_{1F,P}(x_2) = x_2^{*\prime}\gamma_{1F}$ and $h_{2F,P}(x_2) = x_2^{*\prime}\gamma_{2F}$ with $\gamma_{1F} = [\mathbb{E}_F(X_2^* X_2^{*\prime})]^{-1}\mathbb{E}_F(X_2^* Y)$ and $\gamma_{2F} = [\mathbb{E}_F(X_2^* X_2^{*\prime})]^{-1}\mathbb{E}_F(X_2^* X_1')$, which can then be plugged into (D.1) to obtain the influence function of $\hat{\beta}_{n,P}$. In order to get its sample version, let $\hat{\gamma}_{1,n} \equiv \left(\sum_{i=1}^n X_{2,i}^* X_{2,i}^{*\prime}\right)^{-1}\left(\sum_{i=1}^n X_{2,i}^* Y_i\right)$ and $\hat{\gamma}_{2,n} \equiv \left(\sum_{i=1}^n X_{2,i}^* X_{2,i}^{*\prime}\right)^{-1}\left(\sum_{i=1}^n X_{2,i}^* X_{1,i}'\right)$, then one has

$$\psi_{n,P}(Z_i) = -\left[\frac{1}{n}\sum_{i=1}^n (X_{1,i} - X_{2,i}^{*\prime}\hat{\gamma}_{2,n})(X_{1,i} - X_{2,i}^{*\prime}\hat{\gamma}_{2,n})'\right]^{-1} \cdot \{Y_i - X_{2,i}^{*\prime}\hat{\gamma}_{1,n} - (X_{1,i} - X_{2,i}^{*\prime}\hat{\gamma}_{2,n})'\hat{\beta}_{n,SP}\} \cdot (X_{1,i} - X_{2,i}^{*\prime}\hat{\gamma}_{2,n}), \tag{D.2}$$

where note that $\beta$ in the influence function (D.1) is replaced by its robust estimator $\hat{\beta}_{n,SP}$.

For the semiparametric estimator, a series of basis functions $G^L(x_1, x_2) \equiv (g_{1L}(x_1, x_2), g_{2L}(x_1, x_2), \ldots, g_{LL}(x_1, x_2))'$ is used to approximate the unknown function $s(x_1, x_2)$ in the original model, where $L$ is an integer that increases with $n$ and $g_{lL}(x_1, x_2)$ is a known function (e.g., polynomial functions) for each $l \in \{1, \ldots, L\}$. In this case, $\hat{\beta}_{n,SP}$ is just the least squares coefficient of $X_1$ in a linear regression of $Y$ on $X_1$ and $G^L(X_1, X_2)$, and its influence function is what is in (D.1). The argument in Ackerberg, Chen and Hahn (2012) and Ackerberg, Chen, Hahn and Liao (2014) allows one to treat this series approximation as the true model in estimating the asymptotic variance of $\hat{\beta}_{n,SP}$. To proceed, let

$$\hat{\lambda}_{1,n} \equiv \left(\sum_{i=1}^n G^L(X_{1,i}, X_{2,i}) G^{L\prime}(X_{1,i}, X_{2,i})\right)^{-1}\left(\sum_{i=1}^n G^L(X_{1,i}, X_{2,i}) Y_i\right), \text{ and}$$
$$\hat{\lambda}_{2,n} \equiv \left(\sum_{i=1}^n G^L(X_{1,i}, X_{2,i}) G^{L\prime}(X_{1,i}, X_{2,i})\right)^{-1}\left(\sum_{i=1}^n G^L(X_{1,i}, X_{2,i}) X_{1,i}'\right),$$

then one has

$$\psi_{n,SP}(Z_i) = -\left[\frac{1}{n}\sum_{i=1}^n (X_{1,i} - G^{L\prime}(X_{1,i}, X_{2,i})\hat{\lambda}_{2,n})(X_{1,i} - G^{L\prime}(X_{1,i}, X_{2,i})\hat{\lambda}_{2,n})'\right]^{-1} \cdot \{Y_i - G^{L\prime}(X_{1,i}, X_{2,i})\hat{\lambda}_{1,n} - (X_{1,i} - G^{L\prime}(X_{1,i}, X_{2,i})\hat{\lambda}_{2,n})'\hat{\beta}_{n,SP}\} \cdot (X_{1,i} - G^{L\prime}(X_{1,i}, X_{2,i})\hat{\lambda}_{2,n}). \tag{D.3}$$

As a result, the averaging weight can be constructed by first plugging (D.2) and (D.3) into (C.1), (C.2) and (C.3), and then plugging the latter into (2.2).

Two points are worth emphasizing here. First, $\beta$ naturally arises in the influence functions (D.1) and is invariant to how the conditional mean function $h$ is modeled. Therefore, when computing the sample analogs of the influence functions using (D.2) and (D.3), $\beta$ should be replaced by $\hat{\beta}_{n,SP}$, the estimator that is consistent regardless of whether the joint normality restriction is correctly specified or not. Second, the nuisance function $h$ directly enters the influence function of $\hat{\beta}_n$. As a result, how $h$ is modeled (by the linear function of $x_2$ only or without such restriction) affects the functional form of the influence functions, even though neither (D.2) nor (D.3) contains a correction term for the first step estimation error of $h$.

**Primitive conditions.** For a specific model and specific estimators $\hat{\beta}_{n,SP}$ and $\hat{\beta}_{n,P}$, Conditions 1 and 3 are straightforward to verify, and Condition 2 can be verified under more primitive conditions. The following condition is the primitive condition of Condition 2 for the partially linear model 4.

**Condition 2′.** *Let $\|\cdot\|$ indicate the Euclidean norm of a vector and let $\|\cdot\|_{\mathcal{L}_2}$ indicate the $\mathcal{L}_2$ norm of a function. For the partially linear model in (4.1) and the estimators described in Section 2, assume that the following conditions hold for any $F \in \mathcal{F}$, where $0 < M < \infty$ and $\tau > 0$ are some generic constants.*

*(i)* $\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|s(X_1,X_2))] \cdot [X_1 - \mathbb{E}_F(X_1|s(X_1,X_2))]'\}$ *is positive definite.*

*(ii)* $\mathbb{E}_F\{[U + s(X_1,X_2) - \mathbb{E}_F(s(X_1,X_2)|X_2)]^2[X_1 - \mathbb{E}_F(X_1|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|X_2)]'\}$ *is positive definite.*

*(iii) For functions $s(x_1,x_2)$ and $h_{2F,P}(x_2) \equiv \mathbb{E}_F(X_1|X_2) = x_2^{*\prime}\gamma_{2F}$, there exist $d_s$, $d_2$, $\pi_{s,L}$ and $\pi_{2,L}$ such that $\|s(x_1,x_2) - G^{L\prime}(x_1,x_2)\pi_{s,L}\|_{\mathcal{L}_2} = O(L^{-d_s})$ and $\|h_{2F}(x_2) - G^{L\prime}(x_2)\pi_{2,L}\|_{\mathcal{L}_2} = O(L^{-d_2})$ as $L \to \infty$, where $G^L(x_1,x_2)$ and $G^L(x_2)$ are series basis functions of order $L$.*

*(iv) $var_F[Y|X_1, s(X_1,X_2)] \le M < \infty$ and $var_F[X_1|s(X_1,X_2)] \le M$.*

*(v)* $\mathbb{E}_F\{\|X_1 - \mathbb{E}_F[X_1|s(X_1,X_2)]\|^{2+\tau}\} \le M$.

*(vi) The series order $L$ is such that $L \to \infty$, $L/n \to 0$ and $\sqrt{n}L^{-d_s-d_2} \to 0$ as $n \to 0$.*

*(vii) The variance-covariance matrix (under $F$) of $(X_1', X_2')'$ is positive definite.*

*(viii)* $\mathbb{E}_F[\|X_1\|^{2+\tau}] \le M$ *and* $\mathbb{E}_F[\|X_2\|^{2+\tau}] \le M$.

Condition 2′(i) - (vi) follow Assumption 2 and Theorem 2 in Donald and Newey (1994), which ensure the asymptotic normality of $\hat{\beta}_{n,SP}$ based on series first step. Among them, (i) is the key identification requirement of $\beta_F$; (i) and (ii) ensure that the asymptotic variance-covariance matrix of $\hat{\beta}_{n,SP}$ is well defined; (iii) implies that the nuisance functions can be approximated well by the series basis functions; (iv) implies that they can be consistently estimated; (v) is the moment condition required by the central limit theorem; and (vi) gives the under-smoothing order of the series basis functions. (vii) is the key identification requirement of $\beta_{F,P}$, and (viii) is the usual moment condition for the asymptotic normality of $\hat{\beta}_{n,P}$ based on linear regression of $Y$ on $(X_1', X_2')'$ and an intercept.[5]

**Verification of the primitive conditions.** This part verifies Conditions 1, 2′ and 3 for the Monte Carlo model (4.2) and the estimators used in Section 4. Recall that in this model,

$$s(x_1,x_2) \equiv x_2'\theta_1 + t(x_1,x_2), \text{ with } t(x_1,x_2) \equiv \rho\left(\sum_{j=1}^{4}\theta_{2j}\exp(x_{2j}) + \sum_{j=1}^{4}\theta_{3j}x_{1j}x_{2j}\right). \tag{D.4}$$

Because the misspecified model (4.3) only uses $\theta_0 + x_2'\theta_1$ as $s(x_1,x_2)$, $\rho$ controls the degree of misspecification. Note that for $\forall F \in \mathcal{F}$, one has

$$\begin{bmatrix} \theta_{0,F,P} - \theta_{0,F} \\ \beta_{F,P} - \beta_F \\ \theta_{1,F,P} - \theta_{1,F} \end{bmatrix} = \begin{bmatrix} 1 & \mathbb{E}_F(X_1') & \mathbb{E}_F(X_2') \\ \mathbb{E}_F(X_1) & \mathbb{E}_F(X_1X_1') & \mathbb{E}_F(X_1X_2') \\ \mathbb{E}_F(X_2) & \mathbb{E}_F(X_2X_1') & \mathbb{E}_F(X_2X_2') \end{bmatrix}^{-1} \begin{bmatrix} \mathbb{E}_F[t(X_1,X_2)] \\ \mathbb{E}_F[X_1 t(X_1,X_2)] \\ \mathbb{E}_F[X_2 t(X_1,X_2)] \end{bmatrix}, \tag{D.5}$$

where $(\theta_{0,F,P}', \beta_{F,P}', \theta_{1,F,P}')'$ is the pseudo-true parameter value in the misspecified model (4.3) and $(\theta_{0,F}', \beta_F', \theta_{1,F}')'$ is the true parameter vector in model (4.2).[6] The joint normal distribution of $(X_1', X_2')'$ implies that on the right hand side of (D.5), the entries of the first matrix are non-zero finite numbers, and the second vector is proportional to $\rho$, since $\mathbb{E}_F(X_l X_{1j} X_{2j})$ and $\mathbb{E}_F[X_l \exp(X_{2j})]$ are both finite ($l = 1, 2$ and

[5] The joint asymptotic normality can be shown under Condition 2′ by invoking Cramér-Wold theorem (not elaborated here).
[6] In particular, $\theta_{0,F} = 0$.

$j = 1, 2, 3, 4$).[7] So, one gets

$$\delta_F \equiv \beta_{F,P} - \beta_F = c_1\rho \tag{D.6}$$

with non-zero $c_1$, where the non-zero constant $c_1$ depends on the moments of polynomials up to the third order and the exponential functions of $(X_1', X_2')'$. As a result, as long as the values of $\rho$ contain an open set around 0, Condition 1(ii) is satisfied. Moreover, note that the nuisance function $h_F$ in Condition 1(i) is $s(x_1, x_2)$ and $g_{\gamma_F}$ in Condition 1(i) is $x_2'\theta_{1,F,P}$, then one has

$$\|g_{\gamma_F} - h_F\|_{\mathcal{L}_2} = \|x_2'(\theta_{1,F,P} - \theta_{1,F}) + \theta_{0,F,P} - t(x_1, x_2)\|_{\mathcal{L}_2} = c_2|\rho|, \text{ for some } c_2 > 0,$$

where the second equality holds due to (D.5), the definition of $t(x_1, x_2)$ in (D.4), the joint normal distribution of $(X_1', X_2')'$ and that $\theta_{0,F} = 0$ in model (4.2). As a result, Condition 1(i) is satisfied.

The joint normal distribution of $(X_1', X_2')'$ in Section 4 immediately implies that Conditions $2'$(vii), (viii) and the second part of (iv) are satisfied.[8] Moreover, the normal distribution of $U$ and its independence from $(X_1', X_2')'$ ensure Condition $2'$ (ii) and the first part of (iv). In addition, the definition of $s(x_1, x_2)$ in (D.4) shows that $X_1$ and $s(X_1, X_2)$ are not perfectly collinear, indicating that Condition $2'$(i) and (v) are satisfied. Furthermore, note that $s(x_1, x_2)$ only contains linear, quadratic and exponential functions of $x_1$ and $x_2$ and $h_{2F,P}(x_2)$ only contains linear function of $x_2$, which are all four times continuously differentiable functions, so Condition $2'$(iii) is satisfied with $d_s = d_2 = \frac{1}{2}$.[9] Given this, one can choose $L$ such that $L \to \infty$, $L/n \to 0$ and $L^2/n \to \infty$ to satisfy Condition $2'$(vi).

To verify Condition 3, first recall $\delta_F = c_1\rho$ with non-zero $c_1$ shown in (D.6), so $\delta_F$ belongs to a compact set as long as $\rho$ does. Second, note that the asymptotic variance-covariance matrices of $\hat\beta_{n,SP}$ and $\hat\beta_{n,P}$ are

$$V_{F,SP} = \sigma_U^2 \cdot \left(\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|s(X_1, X_2))] \cdot [X_1 - \mathbb{E}_F(X_1|s(X_1, X_2))]'\}\right)^{-1}, \tag{D.7}$$

$$\begin{aligned} V_{F,P} = &\left(\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|X_2)]'\}\right)^{-1} \\ &\cdot \left(\mathbb{E}_F\{[U + s(X_1, X_2) - \mathbb{E}_F(s(X_1, X_2)|X_2)]^2[X_1 - \mathbb{E}_F(X_1|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|X_2)]'\}\right) \\ &\cdot \left(\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|X_2)]'\}\right)^{-1}, \end{aligned} \tag{D.8}$$

$$\begin{aligned} C_F = &\left(\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|s(X_1, X_2))] \cdot [X_1 - \mathbb{E}_F(X_1|s(X_1, X_2))]'\}\right)^{-1} \\ &\cdot \left(\mathbb{E}_F\{[U + s(X_1, X_2) - \mathbb{E}_F(s(X_1, X_2)|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|s(X_1, X_2))]'\}\right) \\ &\cdot \left(\mathbb{E}_F\{[X_1 - \mathbb{E}_F(X_1|X_2)] \cdot [X_1 - \mathbb{E}_F(X_1|X_2)]'\}\right)^{-1}, \end{aligned} \tag{D.9}$$

where $\sigma_U^2 \equiv \mathbb{E}_F(U^2)$. Given the specification in (4.2) (reiterated in (D.4)) and the joint normal distribution of $(X_1', X_2')'$ in Section 4, the following points can be verified.

1. $h_{2F,P}(\cdot) = \mathbb{E}_F(X_1|X_2)$ defined before (D.2) is a $4 \times 1$ vector-valued linear function of $x_2$ that does not depend on $\rho$; in particular, its $j$th coordinate is $h_{2F,P,j}(x_2) \equiv \mathbb{E}_F(X_{1j}|X_2 = x_2) = 0.4 + 0.2\sum_{l=1}^4 x_{2,l}$ for $j = 1, 2, 3, 4$.

2. Note that once the values of $\exp(X_{2j})$ and $X_{1j}X_{2j}$ ($j = 1, 2, 3, 4$) are fixed, then so are the values of $X_{1j}$ and $X_{2j}$ ($j = 1, 2, 3, 4$); and vice versa. For this reason, the function $h_{2F}(\cdot) = \mathbb{E}_F(X_1|s(X_1, X_2)) = \mathbb{E}_F(X_1|X_2, X_{11}X_{21}, X_{12}X_{22}, X_{13}X_{23}, X_{14}X_{24})$ defined before (D.1) does not depend on $\rho$, although its functional form is difficult to obtain and hence is omitted here.[10]

[7] The finite moments of $\exp(X_2)$ can be shown using the moment generating function of the normally distributed $X_2$. For example, $\mathbb{E}\{[\exp(X_{2j})]^2\} = \mathbb{E}[\exp(2X_{2j})] = M_{X_{2j}}(2) = \exp(2\mu_{2j} + 2\sigma_{2j}^2) < \infty$, with $\mu_{2j} = 2$ and $\sigma_{2j}^2 = 0.5^2$.

[8] Note that conditional variance is bounded above by unconditional variance.

[9] $s(x_1, x_2)$ is four times continuously differentiable and has eight arguments, so by the discussion that follows Assumption 3 in Newey (1997), $d_s = \frac{4}{8} = \frac{1}{2}$. Similarly, $h_{2F}(x_2)$ is twice continuously differentiable and has four arguments, so $d_2 = \frac{2}{4} = \frac{1}{2}$.

[10] See Footnote 4 for details on $\mathbb{E}_F(X_1|s(X_1, X_2))$.

3. By the specification in (4.2), one has

$$s(X_1, X_2) - \mathbb{E}_F(s(X_1, X_2)|X_2) = \rho \sum_{j=1}^{4} \theta_{3j} X_{2j}[X_{1j} - \mathbb{E}_F(X_{1j}|X_2)] \equiv C_3\rho, \tag{D.10}$$

where $C_3$ is a random variable that depends on $X_1$ and $X_2$.

Point 2 immediately implies that $V_{F,SP}$ in (D.7) equals to

$$V_{F,SP} = \sigma_U^2 \cdot [\mathbb{E}_F(W_{F,SP} W'_{F,SP})]^{-1}, \\ \text{with } W_{F,SP} \equiv X_1 - \mathbb{E}_F(X_1|X_2, X_{11}X_{21}, X_{12}X_{22}, X_{13}X_{23}, X_{14}X_{24}), \tag{D.11}$$

which does not depend on $\rho$. Points 1 - 3 together imply that $V_{F,P}$ in (D.8) equals to

$$V_{F,P} = [\mathbb{E}_F(W_{F,P}W'_{F,P})]^{-1} \cdot \{\mathbb{E}_F[(U + C_6\rho)^2 W_{F,P}W'_{F,P}]\} \cdot [\mathbb{E}_F(W_{F,P}W'_{F,P})]^{-1}, \\ \text{with } W_{F,P} \equiv X_1 - \mathbb{E}_F(X_1|X_2), \tag{D.12}$$

which is a quadratic function of $\rho$. Similarly, one can show that $C_F$ in (D.9) is a linear function of $\rho$. In summary, $\bar{S}(F)$ defined in (3.17) is a quadratic function of $\rho$; that is, for $F \in \mathcal{F}$ such that $\delta_F = c_1\rho$, there exist some fixed vectors $c_4$, $c_5$ and $c_6$ such that $\bar{S}(F) = c_4 + c_5\rho + c_6\rho^2$. This has two implications. First, as long as $\rho$ takes values from a compact set, so does $S(F)$ defined in (3.17), satisfying Condition 3(i). Second, Condition 3(ii) is satisfied with $\kappa = 2$, $\epsilon_F = 1$ and $C$ being some constant depending on $c_4$, $c_5$ and $c_6$ but not $\rho$.

**Verification of $V_{F,SP} \geq V_{F,P}$ in Condition 2(i).** The paragraph that follows Condition 2 provided intuition for $V_{F,SP} \geq V_{F,P}$ using the semiparametric efficiency bound and Le Cam's third lemma. For a specific model, however, this result can often be verified directly. What follows will use (D.11) and (D.12) to verify it for the parameterization of the partially linear model in (4.2).

Recall that (D.6) shows that $\delta_F \equiv \beta_{F,P} - \beta_F = c_1\rho$ with non-zero $c_1$. Also recall that the presumption of Condition 2(i) is $\|d\| < \infty$ with $n^{1/2}\delta_{F_n} \to d$, then any sequence $\rho_n$ of $\rho$ values considered here satisfies $\rho_n = \frac{\delta_{F_n}}{c_1} = O(n^{-1/2})$. Moreover, $U$ is independent of $W_{F,P}$ due to the independence between $U$ and $(X_1', X_2')'$. These together imply that in the scenario of Condition 2(i), $V_{F,P}$ in (D.12) equals to $\sigma_U^2 \cdot [\mathbb{E}_F(W_{F,P}W'_{F,P})]^{-1}$. Comparing it with (D.11), it is obvious that $\mathbb{E}_F(W_{F,SP}W'_{F,SP}) \leq \mathbb{E}_F(W_{F,P}W'_{F,P})$ since $W_{F,SP}$ conditions on more variables. This further implies that $V_{F,SP} \geq V_{F,P}$.

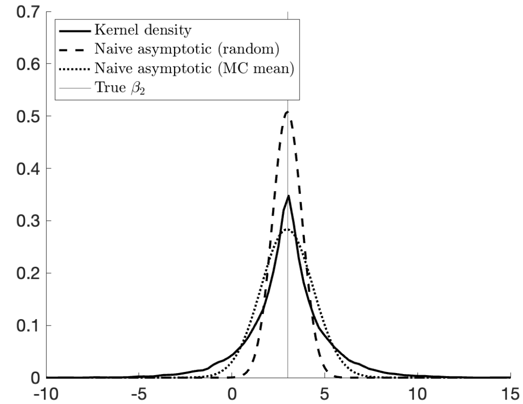
**More Monte Carlo results.** This subsection reports Monte Carlo results for the second, the third and the fourth coordinates of $\beta$ (i.e., $\beta_2$, $\beta_3$ and $\beta_4$). Similar results for $\beta_1$, the first coordinate of $\beta$, are reported in Figure 2 and Table 1 in Section 4.

Figures D.1 - D.3 plot the Monte Carlo distributions (kernel densities) of the averaging estimator $\hat{\beta}_{n,\hat{w}_n}$ (thick solid lines) for representative $\rho$ values. Figure D.1 is for $\beta_2$, Figure D.2 is for $\beta_3$ and Figure D.3 is for $\beta_4$. In the same figures, the normal distributions based on the naive inference method with the common standard error are represented by the thick dashed lines (one randomly chosen Monte Carlo replication) and dotted lines (averaged over all Monte Carlo replications). It is obvious that the naive inference method miscalculates the randomness in the averaging estimators $\hat{\beta}_{n,\hat{w}_n,2}$, $\hat{\beta}_{n,\hat{w}_n,3}$ and $\hat{\beta}_{n,\hat{w}_n,4}$, since it treats the averaging weight $\hat{w}_n$ as non-random.

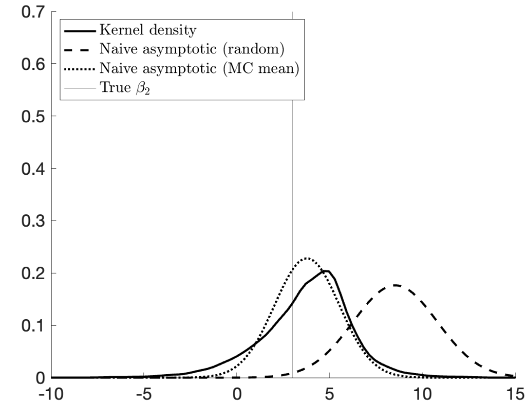
Tables D.1 - D.3 report for different $\rho$ values the rejection rates of $\hat{\beta}_{n,SP}$ with the common standard error and those of $\hat{\beta}_{n,\hat{w}_n}$ with both the naive and the two-step inference methods ($S = 1000$ random draws in the

Figure D.1: True vs. Naive Distributions of $\hat{\beta}_{n,\hat{w}_n,2}$ for the Partially Linear Model

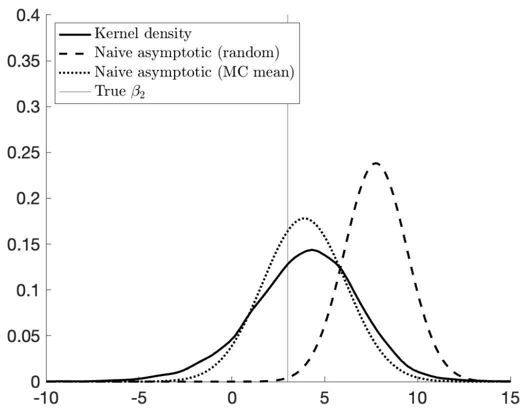
(a) Correct Specification ($\rho = 0$):



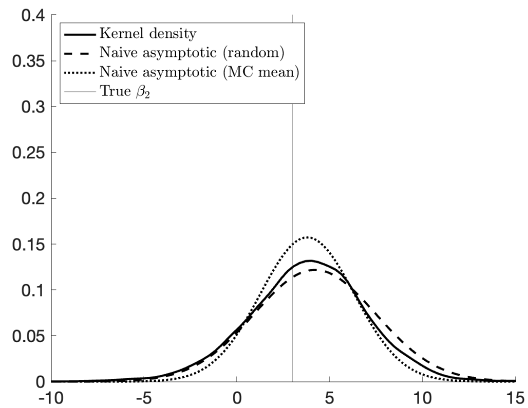

(b) Misspecification ($\rho = 0.2$):



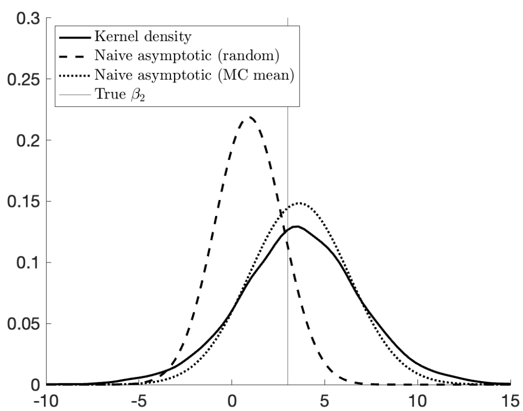

(c) Misspecification ($\rho = 0.4$):



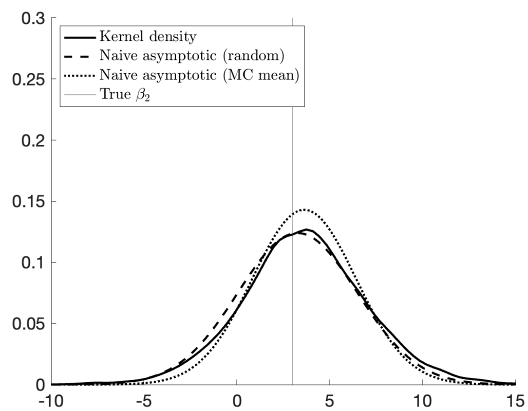

(d) Misspecification ($\rho = 0.6$):




(e) Misspecification ($\rho = 0.8$):



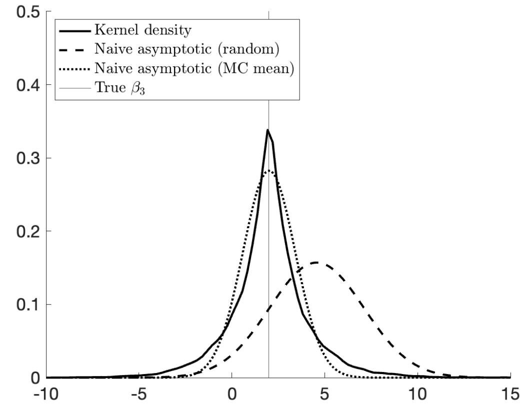

(f) Misspecification ($\rho = 1$):



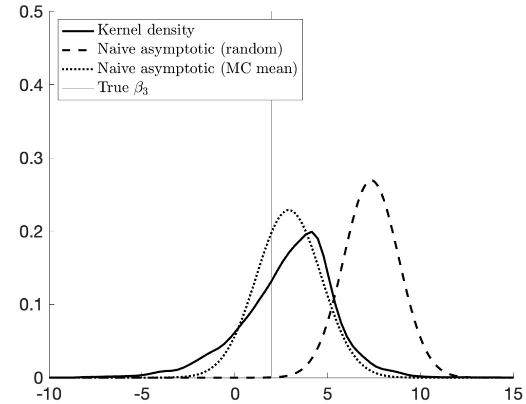

Notes: (1) All distributions are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size.
(2) Solid lines represent the MC distributions of $\hat{\beta}_{n,\hat{w}_n,2}$, the averaging estimator of $\beta_2$. Dashed and dotted lines both represent the asymptotic distribution of $\hat{\beta}_{n,\hat{w}_n,2}$ if the naive inference method, which takes $\hat{w}_n$ as fixed, is used. The former show a randomly chosen MC replication, while the latter show the average over all MC replications.
(3) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

Figure D.2: True vs. Naive Distributions of $\hat{\beta}_{n,\hat{w}_n,3}$ for the Partially Linear Model

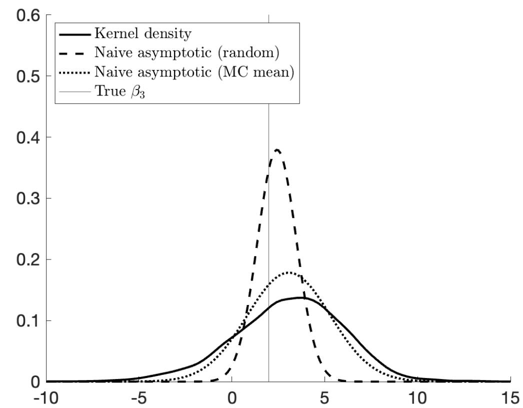
(a) Correct Specification ($\rho = 0$):



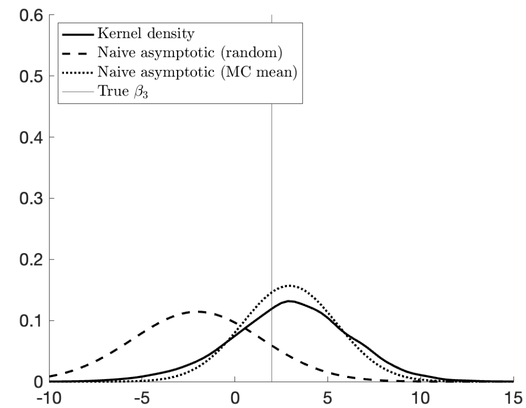

(b) Misspecification ($\rho = 0.2$):



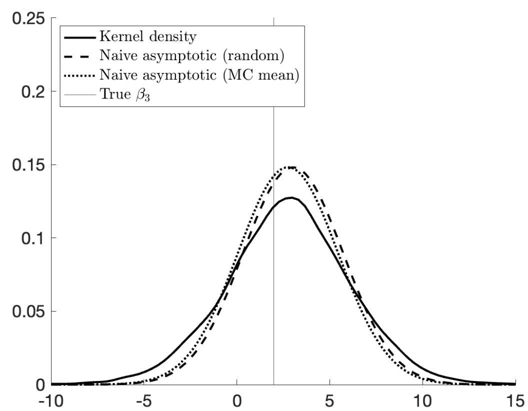

(c) Misspecification ($\rho = 0.4$):



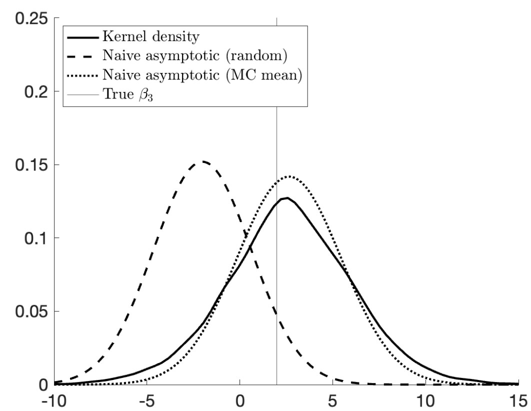

(d) Misspecification ($\rho = 0.6$):




(e) Misspecification ($\rho = 0.8$):



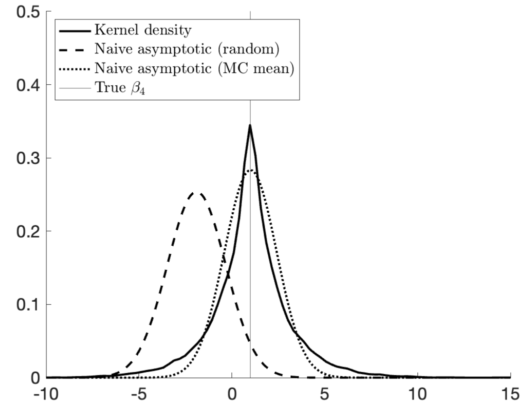

(f) Misspecification ($\rho = 1$):



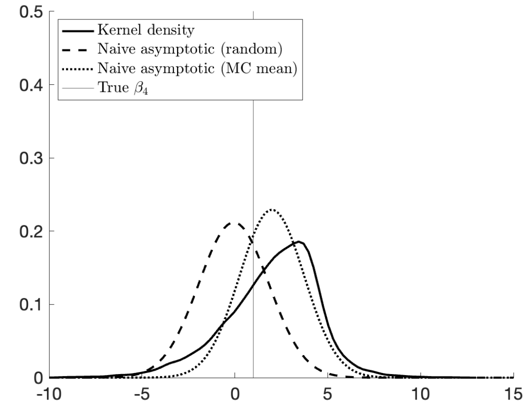

Notes: (1) All distributions are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size.
(2) Solid lines represent the MC distributions of $\hat{\beta}_{n,\hat{w}_n,3}$, the averaging estimator of $\beta_3$. Dashed and dotted lines both represent the asymptotic distribution of $\hat{\beta}_{n,\hat{w}_n,3}$ if the naive inference method, which takes $\hat{w}_n$ as fixed, is used. The former show a randomly chosen MC replication, while the latter show the average over all MC replications.
(3) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

Figure D.3: True vs. Naive Distributions of $\hat{\beta}_{n,\hat{w}_n,4}$ for the Partially Linear Model

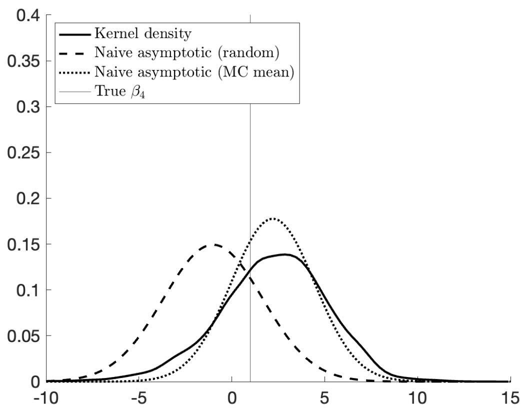
(a) Correct Specification ($\rho = 0$):



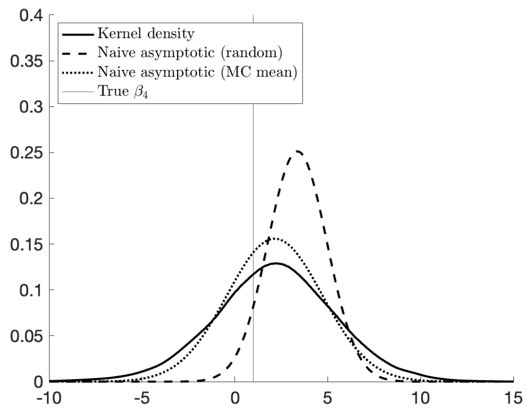

(b) Misspecification ($\rho = 0.2$):



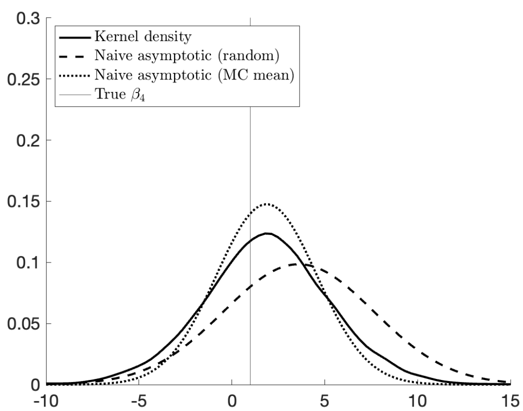

(c) Misspecification ($\rho = 0.4$):



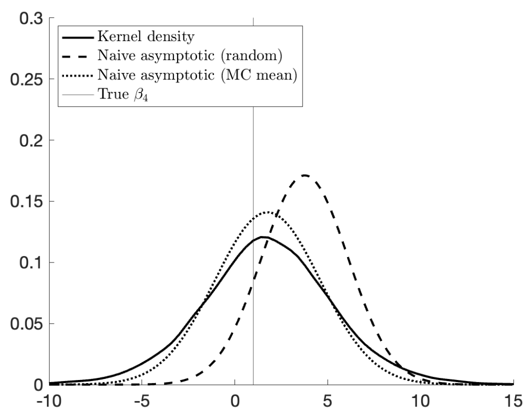

(d) Misspecification ($\rho = 0.6$):




(e) Misspecification ($\rho = 0.8$):




(f) Misspecification ($\rho = 1$):




Notes: (1) All distributions are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size.
(2) Solid lines represent the MC distributions of $\hat{\beta}_{n,\hat{w}_n,4}$, the averaging estimator of $\beta_4$. Dashed and dotted lines both represent the asymptotic distribution of $\hat{\beta}_{n,\hat{w}_n,4}$ if the naive inference method, which takes $\hat{w}_n$ as fixed, is used. The former show a randomly chosen MC replication, while the latter show the average over all MC replications.
(3) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

Table D.1: Rejection Rates for $\hat{\beta}_{n,\hat{w}_n,2}$ in the Partially Linear Model (5% Level)

| $\rho$ | $\hat{\beta}_{n,SP,2}$ | | $\hat{\beta}_{n,\hat{w}_n,2}$ | | | | | | CI Length |
|---|---|---|---|---|---|---|---|---|---|
| | | | Naive | | Naive (robust SE) | | Two-step | | $\frac{CI(\hat{\beta}_{n,\hat{w}_n,2})}{CI(\hat{\beta}_{n,SP,2})}$ |
| | Size | Power | Size | Power | Size | Power | Size | Power | |
| 0.00 | 9.19% | 24.71% | 9.20% | 64.11% | 9.10% | 63.93% | 1.55% | 10.96% | 33.1069 |
| 0.05 | 10.03% | 25.49% | 13.89% | 66.46% | 13.60% | 66.38% | 1.73% | 14.32% | 33.1784 |
| 0.10 | 9.36% | 23.99% | 18.18% | 65.45% | 17.99% | 65.38% | 1.77% | 16.07% | 33.2881 |
| 0.15 | 8.94% | 24.48% | 21.56% | 63.96% | 21.39% | 63.89% | 2.01% | 18.84% | 33.4046 |
| 0.20 | 9.80% | 24.67% | 22.22% | 61.40% | 22.06% | 61.36% | 2.48% | 22.26% | 33.5380 |
| 0.25 | 9.34% | 24.71% | 21.00% | 58.04% | 20.96% | 57.98% | 2.86% | 23.80% | 33.6886 |
| 0.30 | 9.93% | 24.78% | 20.19% | 53.85% | 20.10% | 53.83% | 3.67% | 25.05% | 33.8610 |
| 0.35 | 9.56% | 24.86% | 19.26% | 51.33% | 19.21% | 51.24% | 4.08% | 26.55% | 33.9920 |
| 0.40 | 9.85% | 25.03% | 17.26% | 48.73% | 17.21% | 48.71% | 4.71% | 26.65% | 34.1516 |
| 0.45 | 9.11% | 23.67% | 15.68% | 44.38% | 15.67% | 44.35% | 4.92% | 25.28% | 34.2996 |
| 0.50 | 9.99% | 24.93% | 15.83% | 43.27% | 15.79% | 43.22% | 5.67% | 26.60% | 34.4368 |
| 0.55 | 9.95% | 24.79% | 14.70% | 41.33% | 14.65% | 41.29% | 5.54% | 26.01% | 34.5620 |
| 0.60 | 9.43% | 24.07% | 13.32% | 39.86% | 13.32% | 39.80% | 5.57% | 25.21% | 34.6631 |
| 0.65 | 9.64% | 24.30% | 12.89% | 37.71% | 12.88% | 37.73% | 6.00% | 24.58% | 34.7839 |
| 0.70 | 9.74% | 24.48% | 13.03% | 36.81% | 13.04% | 36.81% | 6.25% | 24.56% | 34.8843 |
| 0.75 | 9.68% | 24.40% | 12.19% | 35.12% | 12.15% | 35.11% | 6.10% | 23.76% | 34.9779 |
| 0.80 | 10.40% | 24.34% | 12.28% | 34.57% | 12.28% | 34.56% | 6.40% | 23.90% | 35.0621 |
| 0.85 | 9.94% | 24.49% | 11.89% | 34.28% | 11.89% | 34.28% | 6.40% | 23.52% | 35.1280 |
| 0.90 | 9.93% | 24.92% | 11.90% | 33.84% | 11.88% | 33.82% | 6.41% | 23.66% | 35.2114 |
| 0.95 | 9.93% | 24.16% | 11.43% | 32.50% | 11.42% | 32.52% | 6.31% | 22.65% | 35.2659 |
| 1.00 | 9.69% | 24.71% | 11.30% | 32.19% | 11.30% | 32.18% | 6.24% | 22.80% | 35.3216 |
| 1.05 | 9.98% | 25.03% | 11.41% | 32.58% | 11.41% | 32.55% | 6.43% | 22.92% | 35.3794 |
| 1.10 | 10.46% | 24.29% | 11.98% | 31.07% | 11.98% | 31.05% | 6.75% | 21.94% | 35.4232 |
| 1.15 | 9.80% | 24.67% | 10.98% | 31.84% | 10.95% | 31.85% | 6.59% | 22.27% | 35.4762 |
| 1.20 | 10.11% | 24.46% | 10.77% | 30.80% | 10.77% | 30.79% | 6.23% | 21.67% | 35.5112 |
| 1.25 | 10.22% | 23.85% | 11.30% | 29.97% | 11.30% | 29.96% | 6.68% | 21.26% | 35.5386 |
| 1.30 | 10.43% | 24.41% | 11.34% | 30.03% | 11.33% | 30.02% | 6.41% | 21.53% | 35.5749 |

Notes: (1) This table reports the inference results for $\hat{\beta}_{n,\hat{w}_n,2}$, the averaging estimator of $\beta_2$.
(2) All results are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size. The two-step inference method uses $S = 1000$ random draws to simulate the distribution of $\bar{\xi}_{F,d} \equiv (1 - w_F)\xi_{F,SP} + w_F(\xi_{F,P} + d)$ in (3.22).
(3) The naive inference methods treat the averaging weight $\hat{w}_n$ as non-random, and hence underestimate the randomness in $\hat{\beta}_{n,\hat{w}_n,2}$. Two naive methods are reported here: the first uses the common estimators of $V_{F,P}$ and $C_F$, which might be biased under misspecification (see the discussion after (C.3)); and the second (robust SE) uses the robust influence function (D.2) when computing the standard error (see Appendix D for details).
(4) The test value for the "Size" columns is 3, the true value of $\beta_2$; the test value for the "Power" columns is 0.
(5) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

Table D.2: Rejection Rates for $\hat{\beta}_{n,\hat{w}_n,3}$ in the Partially Linear Model (5% Level)

| $\rho$ | $\hat{\beta}_{n,SP,3}$ | | $\hat{\beta}_{n,\hat{w}_n,3}$ | | | | | | CI Length |
|---|---|---|---|---|---|---|---|---|---|
| | | | Naive | | Naive (robust SE) | | Two-step | | $\frac{CI(\hat{\beta}_{n,\hat{w}_n,3})}{CI(\hat{\beta}_{n,SP,3})}$ |
| | Size | Power | Size | Power | Size | Power | Size | Power | |
| 0.00 | 9.35% | 16.49% | 9.42% | 46.84% | 9.22% | 46.65% | 1.80% | 6.06% | 31.8093 |
| 0.05 | 9.69% | 16.52% | 14.45% | 52.31% | 14.17% | 52.17% | 1.93% | 8.16% | 31.8198 |
| 0.10 | 9.49% | 16.33% | 20.92% | 53.27% | 20.61% | 53.14% | 2.20% | 10.44% | 31.8614 |
| 0.15 | 10.07% | 16.77% | 25.42% | 52.54% | 25.21% | 52.52% | 2.78% | 13.99% | 31.9477 |
| 0.20 | 9.63% | 17.04% | 25.25% | 50.24% | 25.09% | 50.21% | 3.20% | 17.46% | 31.0517 |
| 0.25 | 9.65% | 17.02% | 23.98% | 47.67% | 23.86% | 47.65% | 4.42% | 20.00% | 32.1615 |
| 0.30 | 10.15% | 16.74% | 22.78% | 43.62% | 22.56% | 43.59% | 5.70% | 20.86% | 32.3306 |
| 0.35 | 9.61% | 16.22% | 20.77% | 40.84% | 20.76% | 40.77% | 6.25% | 21.78% | 32.4561 |
| 0.40 | 9.90% | 16.80% | 19.42% | 38.84% | 19.35% | 38.83% | 7.05% | 23.04% | 32.6202 |
| 0.45 | 10.06% | 16.79% | 18.22% | 35.95% | 18.20% | 35.87% | 7.97% | 22.56% | 32.7834 |
| 0.50 | 10.25% | 15.83% | 16.34% | 32.80% | 16.28% | 32.77% | 7.71% | 21.51% | 32.9379 |
| 0.55 | 9.89% | 16.35% | 15.58% | 31.71% | 15.56% | 31.67% | 8.12% | 21.49% | 33.0857 |
| 0.60 | 10.37% | 16.99% | 15.40% | 30.73% | 15.35% | 30.74% | 8.68% | 21.88% | 33.2150 |
| 0.65 | 9.91% | 16.83% | 14.31% | 29.00% | 14.30% | 29.00% | 8.37% | 20.75% | 33.3506 |
| 0.70 | 10.14% | 17.00% | 13.90% | 28.56% | 13.86% | 28.51% | 8.43% | 20.84% | 33.4695 |
| 0.75 | 9.46% | 16.54% | 12.88% | 26.52% | 12.87% | 26.50% | 8.05% | 19.66% | 33.5899 |
| 0.80 | 10.62% | 17.09% | 13.44% | 26.64% | 13.42% | 26.60% | 8.71% | 20.37% | 33.6846 |
| 0.85 | 9.89% | 16.79% | 12.35% | 25.31% | 12.35% | 25.31% | 8.05% | 19.32% | 33.7756 |
| 0.90 | 10.71% | 17.31% | 12.88% | 25.08% | 12.87% | 25.10% | 8.72% | 19.34% | 33.8567 |
| 0.95 | 10.64% | 16.45% | 12.50% | 23.82% | 12.49% | 23.81% | 8.11% | 18.27% | 33.9313 |
| 1.00 | 10.08% | 16.81% | 12.13% | 23.45% | 12.12% | 23.44% | 8.02% | 18.17% | 33.9944 |
| 1.05 | 10.93% | 16.74% | 12.61% | 23.32% | 12.61% | 23.30% | 8.63% | 18.13% | 34.0518 |
| 1.10 | 10.58% | 17.66% | 12.37% | 23.85% | 12.36% | 23.84% | 8.58% | 18.72% | 34.1105 |
| 1.15 | 10.69% | 17.19% | 12.31% | 22.54% | 12.32% | 22.53% | 8.35% | 18.01% | 34.1666 |
| 1.20 | 10.91% | 17.07% | 12.18% | 22.46% | 12.17% | 22.41% | 8.71% | 17.61% | 34.2091 |
| 1.25 | 11.08% | 17.57% | 12.51% | 22.91% | 12.52% | 22.93% | 8.63% | 17.98% | 34.2507 |
| 1.30 | 11.08% | 17.38% | 12.50% | 21.84% | 12.49% | 21.86% | 8.75% | 17.59% | 34.2906 |

Notes: (1) This table reports the inference results for $\hat{\beta}_{n,\hat{w}_n,3}$, the averaging estimator of $\beta_3$.
(2) All results are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size. The two-step inference method uses $S = 1000$ random draws to simulate the distribution of $\bar{\xi}_{F,d} \equiv (1 - w_F)\xi_{F,SP} + w_F(\xi_{F,P} + d)$ in (3.22).
(3) The naive inference methods treat the averaging weight $\hat{w}_n$ as non-random, and hence underestimate the randomness in $\hat{\beta}_{n,\hat{w}_n,3}$. Two naive methods are reported here: the first uses the common estimators of $V_{F,P}$ and $C_F$, which might be biased under misspecification (see the discussion after (C.3)); and the second (robust SE) uses the robust influence function (D.2) when computing the standard error (see Appendix D for details).
(4) The test value for the "Size" columns is 2, the true value of $\beta_3$; the test value for the "Power" columns is 0.
(5) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

Table D.3: Rejection Rates for $\hat{\beta}_{n,\hat{w}_n,4}$ in the Partially Linear Model (5% Level)

| $\rho$ | $\hat{\beta}_{n,SP,4}$ | | $\hat{\beta}_{n,\hat{w}_n,4}$ | | | | | | CI Length |
|---|---|---|---|---|---|---|---|---|---|
| | | | Naive | | Naive (robust SE) | | Two-step | | $\frac{CI(\hat{\beta}_{n,\hat{w}_n,4})}{CI(\hat{\beta}_{n,SP,4})}$ |
| | Size | Power | Size | Power | Size | Power | Size | Power | |
| 0.00 | 10.25% | 11.89% | 10.34% | 24.59% | 10.19% | 24.24% | 1.61% | 2.57% | 33.1842 |
| 0.05 | 9.75% | 11.86% | 15.86% | 34.95% | 15.67% | 34.70% | 2.00% | 3.43% | 33.1979 |
| 0.10 | 9.60% | 11.69% | 23.23% | 38.97% | 23.00% | 38.78% | 1.97% | 4.47% | 33.2342 |
| 0.15 | 9.61% | 11.45% | 27.28% | 40.30% | 27.13% | 40.29% | 2.58% | 5.88% | 33.2914 |
| 0.20 | 9.26% | 11.05% | 27.23% | 39.53% | 27.16% | 39.41% | 3.19% | 8.01% | 33.3738 |
| 0.25 | 9.52% | 11.28% | 26.18% | 36.56% | 26.08% | 36.48% | 4.38% | 10.66% | 33.4788 |
| 0.30 | 10.48% | 11.81% | 25.09% | 34.94% | 25.09% | 34.85% | 6.25% | 13.17% | 33.6082 |
| 0.35 | 10.01% | 11.59% | 22.91% | 32.41% | 22.87% | 32.35% | 7.43% | 14.00% | 33.7167 |
| 0.40 | 8.75% | 10.72% | 19.98% | 29.64% | 19.96% | 28.59% | 7.45% | 14.08% | 33.8650 |
| 0.45 | 9.77% | 11.76% | 19.52% | 26.84% | 19.54% | 26.77% | 8.63% | 15.06% | 34.0064 |
| 0.50 | 9.53% | 11.68% | 18.34% | 25.77% | 18.33% | 25.71% | 8.81% | 15.00% | 34.1383 |
| 0.55 | 10.01% | 11.73% | 17.10% | 23.62% | 17.09% | 23.57% | 8.96% | 14.76% | 34.2743 |
| 0.60 | 9.84% | 11.79% | 16.38% | 23.08% | 16.38% | 23.08% | 8.82% | 14.94% | 34.3891 |
| 0.65 | 10.58% | 12.45% | 15.89% | 21.82% | 15.86% | 21.77% | 9.62% | 14.52% | 34.5273 |
| 0.70 | 9.89% | 11.50% | 14.72% | 20.47% | 14.67% | 20.49% | 8.77% | 13.84% | 34.6379 |
| 0.75 | 10.26% | 12.13% | 14.46% | 20.10% | 14.45% | 20.11% | 8.69% | 13.84% | 34.7625 |
| 0.80 | 10.48% | 12.44% | 14.12% | 19.43% | 14.11% | 19.44% | 9.18% | 13.83% | 34.8632 |
| 0.85 | 10.07% | 11.54% | 13.32% | 18.49% | 13.30% | 18.47% | 8.36% | 12.68% | 34.9520 |
| 0.90 | 11.21% | 12.84% | 14.21% | 18.65% | 14.20% | 18.65% | 9.07% | 13.43% | 35.0439 |
| 0.95 | 11.84% | 13.21% | 14.45% | 18.67% | 14.43% | 18.66% | 9.36% | 13.47% | 35.1175 |
| 1.00 | 10.83% | 12.16% | 13.06% | 17.49% | 13.03% | 17.49% | 8.21% | 12.65% | 35.1874 |
| 1.05 | 11.01% | 12.42% | 13.15% | 17.24% | 13.15% | 17.25% | 8.79% | 12.42% | 35.2540 |
| 1.10 | 11.90% | 13.29% | 13.68% | 17.32% | 13.66% | 17.32% | 8.80% | 12.65% | 35.3121 |
| 1.15 | 11.26% | 12.32% | 12.76% | 16.22% | 12.77% | 16.22% | 8.39% | 11.92% | 35.3659 |
| 1.20 | 11.22% | 12.93% | 12.98% | 17.44% | 12.99% | 17.43% | 8.58% | 12.29% | 35.4167 |
| 1.25 | 11.12% | 12.97% | 12.85% | 16.68% | 12.84% | 16.67% | 8.78% | 12.54% | 35.4533 |
| 1.30 | 11.85% | 13.16% | 12.97% | 16.79% | 13.00% | 16.79% | 8.77% | 12.14% | 35.4980 |

Notes: (1) This table reports the inference results for $\hat{\beta}_{n,\hat{w}_n,4}$, the averaging estimator of $\beta_4$.
(2) All results are based on $R = 10000$ Monte Carlo replications and $n = 1000$ sample size. The two-step inference method uses $S = 1000$ random draws to simulate the distribution of $\bar{\xi}_{F,d} \equiv (1 - w_F)\xi_{F,SP} + w_F(\xi_{F,P} + d)$ in (3.22).
(3) The naive inference methods treat the averaging weight $\hat{w}_n$ as non-random, and hence underestimate the randomness in $\hat{\beta}_{n,\hat{w}_n,4}$. Two naive methods are reported here: the first uses the common estimators of $V_{F,P}$ and $C_F$, which might be biased under misspecification (see the discussion after (C.3)); and the second (robust SE) uses the robust influence function (D.2) when computing the standard error (see Appendix D for details).
(4) The test value for the "Size" columns is 1, the true value of $\beta_4$; the test value for the "Power" columns is 0.
(5) See Section 4 for the details of the partially linear model example and the Monte Carlo experiments.

second step). Table D.1 is for $\beta_2$, Table D.2 is for $\beta_3$ and Table D.3 is for $\beta_4$. Two variations of the naive inference method for $\hat{\beta}_{n,\hat{w}_n}$ are considered. The "Naive" one uses the common estimators of $V_{F,P}$ and $C_F$ when computing the standard error, but they can be biased under misspecification (see the discussion after (C.3)). The "Naive (robust SE)" one uses the robust influence function (D.2) when computing the standard error (see Appendix D for details). For the "Size" columns, the test value is the true value of the coordinate (i.e., 3 for $\beta_2$, 2 for $\beta_3$ and 1 for $\beta_4$); for the "Power" columns, the test value is 0. Table D.1 - D.3 also report the average ratios between the lengths of the two-step confidence intervals of $\hat{\beta}_{n,\hat{w}_n,j}$ and of the standard confidence intervals of $\hat{\beta}_{n,SP,j}$ $(j = 2, 3, 4)$.

All these results are categorically similar to those for $\beta_1$.

## Appendix E Justification for $V_{SP} \geq V_P$ in Condition 2

### Justification for Condition 2(i)

This section provides rationale of $V_{SP} \geq V_P$ in Condition 2(i) based on the semiparametric efficiency theory and Le Cam's third lemma. (The subscript $F$ is suppressed throughout this subsection for notational simplicity.) What follows is not the proof of Condition 2(i), since Condition 2 is a maintained assumption and can be verified with corresponding primitive conditions for a specific model (like Appendix D for the partially linear model). This subsection merely argues that $V_{SP} \geq V_P$ in Condition 2(i) holds for quite general semiparametric models as it does not require much more than the setup of the semiparametric model.

Consider a set $\mathcal{P}$ consisting of densities $f(z|\beta, \beta_P, h, \eta)$, where $h$ is the nuisance parameter identified by the objective function $R(h)$ in (3.2), $\beta$ is the parameter of interest identified by $h$ and the objective function $Q(\beta, h)$ in (3.1), $\beta_P$ is the parameter identified by $g_\gamma$ and the objective function $Q(\beta, g_\gamma)$ in (3.8),[11] and let $\eta \in \mathcal{E}$ denote the parameter that determines the features of the distribution of $Z$ other than those characterized by $\beta$, $\beta_P$ and $h$.[12] Maintain the assumption that the true density is in $\mathcal{P}$; in other words, $\mathcal{P}$ is the semiparametric model. Let $\underline{V_{SP}}$ and $\underline{V_P}$ denote the efficiency bounds of $\beta$ and $\beta_P$, respectively.

Let $\delta \equiv \beta_P - \beta$, then the densities in $\mathcal{P}$ can be rewritten as $f(z|\beta, \beta+\delta, h, \eta)$. For any $f(z|\beta, \beta+\delta, h, \eta) \in \mathcal{P}$, one can define a parametric model (a subset of $\mathcal{P}$) that incorporates the parametric restriction

$$
\begin{aligned}
\mathcal{P}_{\beta,\delta,\gamma} \equiv \{f(z|\beta, \beta + \delta, g_\gamma, \eta) :\ & \beta, \delta \in \mathbb{R}^k,\ \gamma \in \mathbb{R}^t; \\
& \gamma \text{ is identified by the objective function } R(g_\gamma) \text{ in (3.7)}; \\
& \delta = 0 \text{ only if } h = g_\gamma \text{ for some } \gamma \in \mathbb{R}^t \} .
\end{aligned}
$$

Note that this parametric model internalizes the parametric restriction and Condition 1(i) that the parametric restriction leads to bias if misspecified. Also note that $\mathcal{P}_{\beta,\delta,\gamma}$ may or may not include $f(z|\beta, \beta+\delta, h, \eta)$ itself, depending on whether $h$ admits the parametric restriction $g_\gamma$.

If the density $f(z|\beta, \beta+\delta, h, \eta)$ itself belongs to $\mathcal{P}_{\beta,\delta,\gamma}$ (i.e., $h = g_\gamma$ for some $\gamma \in \mathbb{R}^t$), then the parametric restriction is correctly specified, and $\mathcal{P}_{\beta,\delta,\gamma}$ is a parametric submodel – that is, a parametric model that includes the true DGP – like that defined by Bickel, Klaassen, Ritov and Wellner (1993, Definition 1 on page 46) or Tsiatis (2006, page 59). As a result, one has $\underline{V_{SP}} \geq \underline{V_P}$ by the definition of the semiparametric efficiency bound – that is, the efficiency bound of the semiparametric model is the supremum of efficiency bounds of all parametric submodels – such as equation (2) on page 46 in Bickel et al. (1993) or equation

[11] Recall that $g_\gamma$ is a given parametric function characterized by $\gamma \in \mathbb{R}^t$, which is identified by the objective function $R(g_\gamma)$ in (3.7). In fact, for given $g_\gamma$ function, one can rewrite $f(z|\beta, \beta_P, h, \eta)$ as $f(z|\beta, \beta_P, h, g_\gamma, \eta)$ to make the dependence of $\beta_P$ on $\gamma$ explicit, but $g_\gamma$ is suppressed here for notational simplicity.

[12] This follows the setup in the proof of Lemma 1 in Ackerberg et al. (2014). $\eta$ may have infinite dimension.

(4.16) in Tsiatis (2006). At the same time, the construction of $\mathcal{P}_{\beta,\delta,\gamma}$ dictates that $\delta = 0$ and $\beta = \beta_P$ when $f(z|\beta, \beta+\delta, h, \eta) \in \mathcal{P}_{\beta,\delta,\gamma}$. This implies that the statement $V_{SP} \geq V_P$ in Condition 2(i) is a plausible condition when the parametric restriction is correctly specified.

**Remark 3.** *By the definition of the efficiency bounds, one has $V_{SP} \geq \underline{V_{SP}}$ and $V_P \geq \underline{V_P}$, and either equality holds if the corresponding estimator is efficient. Because the above shows that $\underline{V_{SP}} \geq \underline{V_P}$, the statement $V_{SP} \geq V_P$ in Condition 2(i) only means that $\hat{\beta}_{n,P}$ is at least as efficient as $\hat{\beta}_{n,SP}$, but does not require $\hat{\beta}_{n,P}$ to be efficient in general. For instance, if $\underline{V_{SP}} > \underline{V_P}$ or $V_{SP} > \underline{V_{SP}}$, then there is room between $V_{SP}$ and $\underline{V_P}$ such that it is possible that the asymptotic variance $V_P$ of some inefficient parametric estimator $\hat{\beta}_{n,P}$ satisfies $V_{SP} > V_P$.*[13]

What follows shows that the relationship $V_{SP} \geq V_P$ remains invariant if the parametric restriction deviates from correct specification to mild misspecification. For any fixed density $f(z|, \beta^*, \beta^*, g_{\gamma^*}, \eta^*)$ in $\mathcal{P}_{\beta^*,\delta^*,\gamma^*}$ (i.e., $\delta^* = 0$ by the construction of $\mathcal{P}_{\beta^*,\delta^*,\gamma^*}$), let $P$ denote the resulting probability measure, and let $P_n = P$ be a sequence of such probability measures (same for all $n \in \mathbb{N}$). Note that $P_n$ corresponds to the case where the parametric restriction is correctly specified. For any nuisance function $h$ that does not admit the functional form $g_\gamma$, one has $f(z|\beta^*, \beta^*+\delta, h, \eta^*) \notin \mathcal{P}_{\beta^*,\delta^*,\gamma^*}$ and $\delta \neq 0$ by the construction of $\mathcal{P}_{\beta^*,\delta^*,\gamma^*}$. Inspired by Theorem 7.2 in Van der Vaart (2000), consider a sequence of such nuisance functions, denoted by $h_n$, such that the resulting densities $f(z|\beta^*, \beta^*+\delta_n, h_n, \eta^*)$ satisfy $\delta_n = \frac{d_n}{\sqrt{n}}$, $d_n \to d$ for some $d \in \mathbb{R}^k$ with $\|d\| \in (0, \infty)$ and the corresponding $g_{\gamma_n}$ are $g_{\gamma^*}$. Note that the sequence of $h_n$, by the construction of $\mathcal{P}_{\beta^*,\delta^*,\gamma^*}$, converges to $g_{\gamma^*}$, since the corresponding $\delta_n$ converges to zero. Let $Q_n$ denote the resulting sequence of probability measures, and it corresponds to the case where the parametric restriction is mildly misspecified. Under a technical condition called *differentiable in quadratic mean at* $\beta^*$,[14] the log likelihood ratio between $Q_n$ and $P_n$ admits the following Taylor expansion with respect to $\beta_P$ (i.e., $\beta^* + \delta_n$) around $\beta^*$:

$$\begin{aligned}
\log \prod_{i=1}^{n} \frac{dQ_n}{dP_n} &= \log \prod_{i=1}^{n} \frac{f_n(Z_i|\beta^*, \beta^*+\delta_n, h_n, \eta^*)}{f(Z_i|\beta^*, \beta^*, g_{\gamma^*}, \eta^*)} \\
&= d' \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_{\beta_P}(Z_i) \right) - \frac{1}{2} d' \left( -\frac{1}{n} \sum_{i=1}^{n} \ddot{\ell}_{\beta_P}(Z_i) \right) d + o_p(1),
\end{aligned}$$

where $\dot{\ell}_{\beta_P}(z) \equiv \frac{\partial f(z|\beta^*,\beta^*,g_{\gamma^*},\eta^*)/\partial \beta_P}{f(z|\beta^*,\beta^*,g_{\gamma^*},\eta^*)}$ is the score function with respect to $\beta_P$ under $P_n$ evaluated at $\beta^*$, and $\ddot{\ell}_{\beta_P}(z) \equiv \frac{\partial^2 f(z|\beta^*,\beta^*,g_{\gamma^*},\eta^*)/\partial\beta\partial\beta'}{f(z|\beta^*,\beta^*,g_{\gamma^*},\eta^*)}$ is the corresponding Hessian matrix. Note that $\mathbb{E}_{P_n}[\dot{\ell}_{\beta_P}(Z_i)] = 0$ and $\mathbb{E}_{P_n}[-\ddot{\ell}_{\beta_P}(Z_i)] = \mathcal{I}_{\beta_P}$ (the Fisher information matrix with respect to $\beta_P$). By the central limit theorem and Cramér-Wold theorem, it can be shown that

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n,SP} - \beta) \\ \sqrt{n}(\hat{\beta}_{n,P} - \beta) \\ \log \prod_{i=1}^{n} \frac{dQ_n}{dP_n} \end{pmatrix} \xrightarrow{P_n} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \\ -\frac{1}{2} d' \mathcal{I}_{\beta_P} d \end{pmatrix}, \begin{pmatrix} V_{SP} & C & \tau_{SP} \\ C & V_P & \tau_P \\ \tau'_{SP} & \tau'_P & d' \mathcal{I}_{\beta_P} d \end{pmatrix} \right), \tag{E.1}$$

where the symbol $\xrightarrow{P_n}$ means that the left hand side converges in distribution to the right hand side if $P_n$ is

[13] A well known special case is the inverse probability weighted (IPW) estimator of the average treatment effect (ATE) with series logit propensity score. For the ATE, Hahn (1998) proves that $\underline{V_{SP}} = \underline{V_P}$ under the correct specification of the parametric restriction, and Hirano, Imbens and Ridder (2003) show that the IPW estimator with series logit propensity score satisfies $V_{SP} = \underline{V_{SP}}$. Together, these require that the parametric estimator has to be efficiency for $V_{SP} \geq V_P$ to hold (with equality). The author thanks an anonymous referee for pointing this out.

[14] See (7.1) in Van der Vaart (2000), for example. This assumption is common and maintained for the majority of the models in the M estimation literature.

the true distribution of the data.[15] This fulfills the assumption of Le Cam's third lemma (Example 6.7 in Van der Vaart, 2000), so by this lemma one gets

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n,SP} - \beta) \\ \sqrt{n}(\hat{\beta}_{n,P} - \beta) \end{pmatrix} \xrightarrow{Q_n} \mathcal{N}\left( \begin{pmatrix} \tau_{SP} \\ \tau_P \end{pmatrix}, \begin{pmatrix} V_{SP} & C \\ C & V_P \end{pmatrix} \right). \tag{E.2}$$

That is, Le Cam's third lemma implies that the asymptotic variance-covariance matrices of $\hat{\beta}_{n,SP}$ and $\hat{\beta}_{n,P}$ remain invariant whether the parametric restriction is correctly specified or mildly misspecified. Together with the earlier condition that $V_{SP} \geq V_P$ under the correct specification, it provides the rationale behind $V_{SP} \geq V_P$ in Condition 2(i).

**Remark 4.** *In (E.1) and (E.2), $\tau_{SP} \equiv \mathbb{E}_{P_n}[\psi_{SP}(Z_i)\dot{\ell}_{\beta_P}(Z_i)]d$ and $\tau_P \equiv \mathbb{E}_{P_n}[\psi_P(Z_i)\dot{\ell}_{\beta_P}(Z_i)]d$ by the central limit theorem, where $\psi_{SP}(z)$ and $\psi_P(z)$ are the (centered) influence functions of $\hat{\beta}_{n,SP}$ and $\hat{\beta}_{n,P}$, respectively. Note that $\hat{\beta}_{n,SP}$ is a regular and asymptotically linear (RAL) estimator of $\beta$ but $\hat{\beta}_{n,P}$ is an RAL estimator of $\beta_P$,[16] then by Theorem 4.2 in Tsiatis (2006), one has $\mathbb{E}_{P_n}[\psi_{SP}(Z_i)\dot{\ell}_{\beta_P}(Z_i)] = 0$ and $\mathbb{E}_{P_n}[\psi_P(Z_i)\dot{\ell}_{\beta_P}(Z_i)] = I_k$ (i.e., the $k \times k$ identity matrix), further implying that $\tau_{SP} = 0$ and $\tau_P = d$. This, combined with the above argument for $V_{SP} \geq V_P$, indicates that the joint asymptotic distribution postulated in Condition 2(i) is in fact a general result for the semiparametric model and the estimators considered in this paper.*

### Justification for Condition 2(ii)

Note that the asymptotic properties of the semiparametric estimator $\hat{\beta}_{n,SP}$ do not depend on whether $\|d\| < \infty$ or $\|d\| = \infty$, so one still has $n^{1/2}(\hat{\beta}_{n,SP} - \beta_{F_n}) \xrightarrow{d.} \xi_{F,SP}$ under the same primitive conditions like those for Condition 2(i).

To study the asymptotic properties of the parametric estimator $\hat{\beta}_{n,P}$ when $\|d\| = \infty$, consider two cases: (i) $\delta_{F_n} = o(1)$; and (ii) $\|\delta_{F_n}\| > c$ for some $c > 0$. For case (i), let $\psi_{F,P}(z)$ denote the (centered) influence function of $\hat{\beta}_{n,P}$ under DGP $F$, which is an $O_p(1)$ term, then by the definition of $\beta_{F,P}$ and $\delta$,

$$\begin{aligned} & n^{1/2}(\hat{\beta}_{n,P} - \beta_{F_n,P}) = n^{-1/2}\sum_{i=1}^{n}\psi_{F_n,P}(Z_i) + o_p(1) \\ \Longrightarrow\ & n^{1/2}(\hat{\beta}_{n,P} - \beta_{F_n}) = n^{1/2}\delta_{F_n} + O_p(1). \end{aligned} \tag{E.3}$$

Note that the presumption of Condition 2(ii) is that $\|n^{1/2}\delta_{F_n}\| \to \|d\| = \infty$, then $n\delta_{F_n}'\delta_{F_n} \to \infty$, which together with (E.3) implies that $\|n^{1/2}(\hat{\beta}_{n,P} - \beta_{F_n})\| \xrightarrow{p.} \infty$.

For case (ii), note that $\beta_{F,P}$ is identified in (3.8), then under the same conditions for $\hat{\beta}_{n,SP} = \beta_{F_n} + o_p(1)$, one gets $\hat{\beta}_{n,P} = \beta_{F_n,P} + o_p(1)$.[17] This, combined with the presumption that $\|\delta_{F_n}\| = \|\beta_{F_n,P} - \beta_{F_n}\| > c$, implies that

$$\|n^{1/2}(\hat{\beta}_{n,P} - \beta_{F_n})\| \geq |\|n^{1/2}(\hat{\beta}_{n,P} - \beta_{F_n,P})\| - \|n^{1/2}\delta_{F_n}\|| = \|n^{1/2}\delta_{F_n}\| \cdot (1 + o_p(1)) \xrightarrow{p.} \infty.$$

## References

[15] The symbol $\xrightarrow{Q_n}$ below is understood in a similar way.

[16] Recall that this paper maintains the assumption that $\hat{\beta}_{n,SP}$ and $\hat{\beta}_{n,P}$ are both locally regular estimators, so that their asymptotic properties hinge on the influence functions.

[17] This is a familiar result for pseudo-true parameter values (e.g., Newey and McFadden, 1994, Section 2).

Ackerberg, D., X. Chen, and J. Hahn (2012). A practical asymptotic variance estimator for two-step semi-pametric estimators. *Review of Economics and Statistics* 94(2), 481-498.

Ackerberg, D., X. Chen, J. Hahn, and Z. Liao (2014). Asymptotic efficiency of semiparametric two-step GMM. *Review of Economic Studies* 81(3), 919-943.

Bickel, P. J., C. A. Klaassen, J. Ritov, and J. A. Wellner (1993). Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press Baltimore.

Donald, S. G. and W. K. Newey (1994). Series estimation of semilinear models. *Journal of Multivariate Analysis* 50(1), 30-40.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 66(2), 135-331.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71(4), 1161-1189.

Ichimura, H. and S. Lee (2010). Characterization of the asymptotic distribution of semiparametric M-estimators. *Journal of Econometrics* 159(2), 252-266.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79(1), 147-168.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothetis testing. *Handbook of Econometrics* 4, 2111-2245.

Tsiatis, A. (2006). Semiparametric theory and missing data. Springer Science & Business Media.

Van der Vaart, A. W. (2000). Asymptotic statistics. *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.