Supplementary Materials for "Sufficient Dimension Reduction Meets Two-Sample Regression Estimation"

Masayuki Hirukawa*
Ryukoku University

Di Liu†
Stata Corp

Artem Prokhorov‡
University of Sydney, CEBDA & CIREQ

June 8, 2026

Abstract

The Supplementary Materials contain five supplements. Supplement A provides all the technical proofs of the theorems stated in the main text. Supplement B describes a procedure for estimating the variance of the proposed estimators. Supplement C discusses why the MSII estimator of Hirukawa and Prokhorov (2018) cannot be easily extended to include SDR. Supplement D presents the full set of Monte Carlo simulation results. Supplement E contains a corrigendum of Hirukawa et al. (2023).

*e-mail: hirukawa@econ.ryukoku.ac.jp.
†e-mail: flyingliudi@gmail.com.
‡e-mail: artem.prokhorov@sydney.edu.au.

Supplement A: Technical Proofs

A.1 A Useful Lemma

We begin by stating a lemma about weak uniform convergence rates of the numerator and denominator of the oracle kernel regression estimator (Eq. (12) of the main text). For later use, this lemma also documents uniform weak consistency of two hypothetical, oracle estimators of the first-order derivatives, namely,

$$\tilde{r}_\ell^{(1)}(z) := -\frac{1}{mh_\ell^2}\sum_{j=1}^{m} X_{2\ell,j}^B K^{(1)}\left(\frac{Z_{\ell,j}^B - z}{h_\ell}\right) \text{ and} \tag{A1}$$

$$\tilde{p}_\ell^{(1)}(z) := -\frac{1}{mh_\ell^2}\sum_{j=1}^{m} K^{(1)}\left(\frac{Z_{\ell,j}^B - z}{h_\ell}\right). \tag{A2}$$

The lemma is a minor modification of Theorems 2 and 6 in Hansen (2008) and thus left without a proof.

Lemma A1. *If Assumptions S1, S3, S4, S6, and P2-P4 hold, then*

$$\sup_{z\in\mathbb{Z}_\ell}\left|E\left\{\tilde{q}_\ell^{(k)}(z)\right\} - q_\ell^{(k)}(z)\right| = O(h^\nu), \text{ and}$$

$$\sup_{z\in\mathbb{Z}_\ell}\left|\tilde{q}_\ell^{(k)}(z) - E\left\{\tilde{q}_\ell^{(k)}(z)\right\}\right| = O_p\left(\sqrt{\frac{\log m}{mh^{1+2k}}}\right)$$

for $q_\ell \in \{r_\ell, p_\ell\}$, $k \in \{0, 1\}$ *and* $\ell \in \{1, \ldots, d_2\}$, *as* $m \to \infty$.

An important corollary of Lemma A1 is that

$$\sup_{z\in\mathbb{Z}_\ell}|\tilde{g}_{2\ell}(z) - g_{2\ell}(z)| = O_p\left(h^\nu + \sqrt{\frac{\log m}{mh}}\right). \tag{A3}$$

This lemma, jointly with (A3), is used as a workhorse in subsequent proofs.

A.2 Proof of Theorem 1

It follows from Eqs. (4), (9) and (12) of the main text that

$$\begin{aligned} Y_i^A &= \hat{X}_{g,i}^{A\top}\beta - \left[E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\} - g_2\left(Z_i^A\right)\right]^\top \beta_2 + \epsilon_i^A \\ &\quad - \left[\tilde{g}_2\left(Z_i^A\right) - E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\}\right]^\top \beta_2 - \left\{\hat{g}_2\left(\hat{Z}_i^A\right) - \tilde{g}_2\left(Z_i^A\right)\right\}^\top \beta_2. \end{aligned}$$

Then, by the definition of PILS-SDR (Eq. (5) of the main text) and Lemma A1,

$$\hat{\beta}_{PS} = Q^{-1}_{\hat{X}_g^A \hat{X}_g^A} R_{\hat{X}_g^A Y^A} := \beta - Q^{-1}_{\hat{X}_g^A \hat{X}_g^A} \left(R_1 - R_2 + R_3 + R_4 \right) \ , \tag{A4}$$

where

$$R_1 := \frac{1}{n} \sum_{i=1}^{n} \hat{X}_{g,i}^A \left[E\left\{ \tilde{g}_2 \left(Z_i^A \right) \middle| Z_i^A \right\} - g_2 \left(Z_i^A \right) \right]^\top \beta_2$$

is the bias term due to kernel smoothing,

$$R_2 := \frac{1}{n} \sum_{i=1}^{n} \hat{X}_{g,i}^A \epsilon_i^A$$

is the sampling error,

$$R_3 := \frac{1}{n} \sum_{i=1}^{n} \hat{X}_{g,i}^A \left[\tilde{g}_2 \left(Z_i^A \right) - E\left\{ \tilde{g}_2 \left(Z_i^A \right) \middle| Z_i^A \right\} \right]^\top \beta_2$$

is the estimation error of the link functions, and

$$R_4 := \frac{1}{n} \sum_{i=1}^{n} \hat{X}_{g,i}^A \left\{ \hat{g}_2 \left(\hat{Z}_i^A \right) - \tilde{g}_2 \left(Z_i^A \right) \right\}^\top \beta_2$$

is the estimation error of the index coefficients. Now, it follows from (A3) and Assumption P3 that

$$\left\| \hat{X}_{g,i}^A - X_{g,i}^A \right\| = O_p \left(h^\nu + \sqrt{\frac{\log m}{mh}} \right) = o_p (1) \tag{A5}$$

uniformly on $\mathbb{Z}$. Then, $Q_{\hat{X}_g^A \hat{X}_g^A} = \sum_{i=1}^n X_{g,i}^A X_{g,i}^{A\top}/n + o_p(1) \xrightarrow{p} \Phi_{PS} > 0$. The same conditions also imply that $R_1 = O_p(h^\nu) = o_p\left(n^{-1/2}\right)$. By a central limit theorem (CLT), $R_2 = O_p\left(n^{-1/2}\right)$. As will follow from Lemmatas A4 and A7, each of R_3 and R_4 is $O_p\left(n^{-1/2}\right)$. Therefore,

$$\hat{\beta}_{PS} = \beta + o_p\left(n^{-1/2}\right) + O_p\left(n^{-1/2}\right) + O_p\left(n^{-1/2}\right) \xrightarrow{p} \beta. \ \blacksquare$$

A.3 Proof of Theorem 2

The proof requires the following lemmata.

Lemma A2. *If Assumptions G1, S1, S3, S4, S6, and P2-P4 hold, then R_3 defined in the proof of Theorem 1 can be approximated by $R_3 = R_3^* + o_p\left(n^{-1/2}\right)$, where*

$$R_3^* := \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} X_{g,i}^A \sum_{\ell=1}^{d_2} \frac{\eta_{2\ell,j}^B K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell} \right) \beta_{2\ell}}{h_\ell p_\ell \left(Z_{\ell,i}^A \right)},$$

and the $o_p\left(n^{-1/2}\right)$ rate is uniform on $\mathbb{Z}$.

Lemma A3. *Let* $\{V_i\}_{i=1}^{n}$ *and* $\{W_j\}_{j=1}^{m}$ *be independent random samples drawn from two distinct populations. Suppose that given these samples, we construct a two-sample U-statistic*

$$U := \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\phi\left(V_i; W_j\right)$$

with a known kernel function $\phi(\cdot;\cdot) \in \mathbb{R}^p$. *Also define* $\theta := E\{\phi(V;W)\} = E(U)$, $\phi_{10}(V) := E\{\phi(V;W)|V\}$ *and* $\phi_{01}(W) := E\{\phi(V;W)|W\}$. *Then,* $\sqrt{n}(U-\theta) \xrightarrow{d} N\left(0_{p\times 1}, \Sigma_{\phi(V)} + \kappa\Sigma_{\phi(W)}\right)$ *as* $n, m \to \infty$ *so that* $n/m \to \kappa \in (0,\infty)$, *where* $\Sigma_{\phi(V)} := Var\{\phi_{10}(V)\}$ *and* $\Sigma_{\phi(W)} := Var\{\phi_{01}(W)\}$.

Lemma A4. *If Assumptions G1, S1, S3, S4, S6, and P2-P4 hold, then* R_3 *defined in the proof of Theorem 1 admits an asymptotic linear representation*

$$\sqrt{n}R_3 = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\sqrt{\kappa}\psi_{21j} + o_p(1) = \frac{\sqrt{\kappa}}{\sqrt{m}}\sum_{j=1}^{m}\sum_{\ell=1}^{d_2} G_\ell^A\left(Z_{\ell,j}^B\right)\eta_{2\ell,j}^B\beta_{2\ell} + o_p(1)$$

as $n, m \to \infty$ *so that* $n/m \to \kappa \in (0,\infty)$.

Lemma A5. *If Assumptions G1, S1-S7 and P2-P4 hold, then*

$$\begin{aligned}&K\left(\frac{\hat{Z}_{\ell,j}^B - \hat{Z}_{\ell,i}^A}{h_\ell}\right) - K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\\ &= K^{(1)}\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\frac{1}{h_\ell}\left(X_{3j}^B - X_{3i}^A\right)^\top\left(\hat{\theta}_\ell - \theta_\ell\right) + O_p\left(\frac{1}{mh^2}\right)\end{aligned}$$

for all $\ell \in \{1,\dots,d_2\}$, *as* $m \to \infty$, *where the* $O_p\left\{\left(mh^2\right)^{-1}\right\}$ *rate is uniform on* $\mathbb{Z}_\ell$.

Lemma A6. *If Assumptions G1, S1-S7 and P2-P4 hold, then*

$$\sqrt{n}\left\{\hat{g}_{2\ell}\left(\hat{Z}_{\ell,i}^A\right) - \tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)\right\} = \sqrt{\kappa}\left\{X_3^A - g_{3\ell}^B\left(Z_\ell^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\sqrt{m}\left(\hat{\theta}_\ell - \theta_\ell\right) + o_p(1)$$

for all $\ell \in \{1,\dots,d_2\}$, *as* $n, m \to \infty$ *so that* $n/m \to \kappa \in (0,\infty)$, *where the* $o_p(1)$ *rate is uniform on* $\mathbb{Z}_\ell$.

Lemma A7. *If Assumptions G1, S1-S7 and P2-P4 hold, then* R_4 *defined in the proof of Theorem 1 admits an asymptotic linear representation*

$$\begin{aligned}\sqrt{n}R_4 &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\sqrt{\kappa}\psi_{22j} + o_p(1)\\ &= \frac{\sqrt{\kappa}}{\sqrt{m}}\sum_{j=1}^{m}\sum_{\ell=1}^{d_2} E\left[X_g^A\left\{X_3^A - g_{3\ell}^B\left(Z_\ell^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_\ell^A\right)\right]\varphi_{\ell,j}^B\left(\theta_\ell\right)\beta_{2\ell} + o_p(1)\end{aligned}$$

as $n, m \to \infty$ so that $n/m \to \kappa \in (0, \infty)$.

A.3.1 Proof of Lemma A2

Before proceeding, we decompose $\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)$ into two parts, i.e., $\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right) := \tilde{M}_{\ell,i}^A + \tilde{H}_{\ell,i}^A$, where

$$\tilde{M}_{\ell,i}^A := \frac{\frac{1}{mh_\ell}\sum_{j=1}^m \tilde{g}_{2\ell}\left(Z_{\ell,j}^B\right) K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)}{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)}, \text{ and}$$
$$\tilde{H}_{\ell,i}^A := \frac{\frac{1}{mh_\ell}\sum_{j=1}^m \eta_{2\ell,j}^B K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)}{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)}.$$

Also let $\tilde{M}_i^A := \left(\tilde{M}_{1,i}^A, \ldots, \tilde{M}_{d_2,i}^A\right)^\top$ and $\tilde{H}_i^A := \left(\tilde{H}_{1,i}^A, \ldots, \tilde{H}_{d_2,i}^A\right)^\top$. Then,

$$\begin{aligned}
R_3 &= \frac{1}{n}\sum_{i=1}^n X_{g,i}^A \tilde{H}_i^{A\top}\beta_2 \\
&+ \frac{1}{n}\sum_{i=1}^n X_{g,i}^A \left[\tilde{M}_i^A - E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\}\right]^\top \beta_2 \\
&+ \frac{1}{n}\sum_{i=1}^n \left(\hat{X}_{g,i}^A - X_{g,i}^A\right)\left[\tilde{g}_2\left(Z_i^A\right) - E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\}\right]^\top \beta_2 \\
&=: R_{31} + R_{32} + R_{33} \ \text{(say)}.
\end{aligned}$$

We start from working on R_{32} and R_{33}. Repeatedly applying Lemma A1 yields

$$\begin{aligned}
\tilde{M}_i^A &= g_2\left(Z_i^A\right) + O_p\left(h^\nu + \sqrt{\frac{\log m}{mh}}\right) = g_2\left(Z_i^A\right) + o_p(1), \text{ and} \\
E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\} &= g_2\left(Z_i^A\right) + O\left(h^\nu\right) = g_2\left(Z_i^A\right) + o(1),
\end{aligned}$$

where the $o_p(1)$ and $o(1)$ rates are uniform on $\mathbb{Z}$. Then, by Assumptions G1 and S1, (A5) and the Cauchy-Schwarz inequality,

$$\begin{aligned}
E\left\|\sqrt{n}R_{32}\right\|^2 &\le \left\{E\left\|X_{g,i}^A\right\|^4\right\}^{1/2}\left\{E\left\|\tilde{M}_i^A - E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\}\right\|^4\right\}^{1/2}\left\|\beta_2\right\|^2 \\
&\le O(1) \times o(1) \times O(1) = o(1), \text{ and} \\
E\left\|\sqrt{n}R_{33}\right\|^2 &\le \left\{E\left\|\hat{X}_{g,i}^A - X_{g,i}^A\right\|^4\right\}^{1/2}\left\{E\left\|\tilde{g}_2\left(Z_i^A\right) - E\left\{\tilde{g}_2\left(Z_i^A\right)\middle| Z_i^A\right\}\right\|^4\right\}^{1/2}\left\|\beta_2\right\|^2 \\
&\le o(1) \times o(1) \times O(1) = o(1).
\end{aligned}$$

Therefore, $E\left(R_3 - R_{31}\right)^2 = o\left(n^{-1}\right)$, or $R_3 = R_{31} + o_p\left(n^{-1/2}\right)$, where the $o_p\left(n^{-1/2}\right)$ rate is uniform on $\mathbb{Z}$.

Now, $\tilde{H}^A_{\ell,i}$ can be rearranged to

$$\tilde{H}^A_{\ell,i} = \frac{1}{mh_\ell}\sum_{j=1}^{m}\eta^B_{2\ell,j}K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\left[\frac{1}{p_\ell\left(Z^A_{\ell,i}\right)}+\left\{\frac{1}{\tilde{p}_\ell\left(Z^A_{\ell,i}\right)}-\frac{1}{p_\ell\left(Z^A_{\ell,i}\right)}\right\}\right].$$

It follows that $R_{31} := R^*_3 + R_{312}$ (say), where

$$R_{312} := \frac{1}{n}\sum_{i=1}^{n}X^A_{g,i}\left\{\frac{1}{\tilde{p}_\ell\left(Z^A_{\ell,i}\right)}-\frac{1}{p_\ell\left(Z^A_{\ell,i}\right)}\right\}\frac{1}{m}\sum_{j=1}^{m}\sum_{\ell=1}^{d_2}\frac{\eta^B_{2\ell,j}K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\beta_{2\ell}}{h_\ell}.$$

Observe that

$$\begin{aligned}\left|\sqrt{n}R_{312}\right| &\le \frac{1}{n}\sum_{i=1}^{n}\left\|X^A_{g,i}\right\|\left\{\sup_{z\in\mathbb{Z}_\ell}\left|\frac{1}{\tilde{p}_\ell(z)}-\frac{1}{p_\ell(z)}\right|\right\}\\ &\quad\times\sqrt{\frac{n}{m}}\sum_{\ell=1}^{d_2}h_\ell^{-1/2}\sum_{j=1}^{m}\frac{\left|\eta^B_{2\ell,j}\right|\left|K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right|\left|\beta_{2\ell}\right|}{\sqrt{mh_\ell}},\end{aligned}$$

where

$$\begin{aligned}&E\left[\left\{\sum_{j=1}^{m}\frac{\left|\eta^B_{2\ell,j}\right|\left|K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right|\left|\beta_{2\ell}\right|}{\sqrt{mh_\ell}}\right\}^2\middle| Z^A_{\ell,i}\right]\\ &=\frac{\beta^2_{2\ell}}{h_\ell}E\left\{\left(\eta^B_{2\ell,j}\right)^2K^2\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\middle| Z^A_{\ell,i}\right\}\\ &=\frac{\beta^2_{2\ell}}{h_\ell}E\left[E\left\{\left(\eta^B_{2\ell,j}\right)^2\middle| Z^B_{\ell,j},Z^A_{\ell,i}\right\}K^2\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\middle| Z^A_{\ell,i}\right].\end{aligned}\tag{A6}$$

By $Z^B_{\ell,j} \perp\!\!\!\perp Z^A_{\ell,i}$, we have $E\left\{\left(\eta^B_{2\ell,j}\right)^2\middle| Z^B_{\ell,j},Z^A_{\ell,i}\right\} = E\left\{\left(\eta^B_{2\ell,j}\right)^2\middle| Z^B_{\ell,j}\right\} =: \sigma^2_{2\ell}\left(Z^B_{\ell,j}\right)$. Then, the right-hand side of (A6) reduces to

$$\begin{aligned}&\frac{\beta^2_{2\ell}}{h_\ell}\int\sigma^2_{2\ell}(t)K^2\left(\frac{t-Z^A_{\ell,i}}{h_\ell}\right)p_\ell(t)\,dt\\ &=\beta^2_{2\ell}\int\sigma^2_{2\ell}\left(Z^A_{\ell,i}+h_\ell u\right)K^2(u)\,p_\ell\left(Z^A_{\ell,i}+h_\ell u\right)du\\ &=\beta^2_{2\ell}\left\{R(K)\,\sigma^2_{2\ell}\left(Z^A_{\ell,i}\right)p_\ell\left(Z^A_{\ell,i}\right)+O\left(h^\nu\right)\right\}=O(1)\end{aligned}$$

by a change of variable $\left(t-Z^A_{\ell,i}\right)/h_\ell =: u$ and $R(K) := \int K^2(u)\,du$. As a result,

$$\begin{aligned}\left|\sqrt{n}R_{312}\right| &\le O_p(1)\times O_p\left(h^\nu+\sqrt{\frac{\log m}{mh}}\right)\times O(1)\times O\left(h^{-1/2}\right)\times O_p(1)\\ &=O_p\left(h^{\nu-1/2}+\sqrt{\frac{\log m}{mh^2}}\right)=o_p(1)\end{aligned}$$

by Lemma A1 and Assumption P3. This completes the proof. ■

A.3.2 Proof of Lemma A3

Let $N := n+m$ and $T_N := \sqrt{N}\,(U-\theta)$. Then, by Theorem 6.1.4 of Lehmann (1999), $T_N = T_N^* + o_p(1)$ holds, where

$$T_N^* := \sqrt{N}\left[\frac{1}{n}\sum_{i=1}^{n}\{\phi_{10}(V_i)-\theta\} + \frac{1}{m}\sum_{j=1}^{m}\{\phi_{01}(W_j)-\theta\}\right]$$

is asymptotically normal. More specifically, it follows from $\sum_{i=1}^{n}\{\phi_{10}(V_i)-\theta\}/\sqrt{n} \xrightarrow{d} N\left(0_{p\times 1}, \Sigma_{\phi(V)}\right)$, $\sum_{j=1}^{m}\{\phi_{01}(W_j)-\theta\}/\sqrt{m} \xrightarrow{d} N\left(0_{p\times 1}, \Sigma_{\phi(W)}\right)$ and $\phi_{10}(V_i) \perp\!\!\!\perp \phi_{01}(W_j)$ that

$$\begin{aligned}
T_N^* &= \sqrt{\frac{N}{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\phi_{10}(V_i)-\theta\} + \sqrt{\frac{N}{m}}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\{\phi_{01}(W_j)-\theta\} \\
&= \sqrt{\frac{1+\kappa}{\kappa}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\phi_{10}(V_i)-\theta\} + \sqrt{1+\kappa}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\{\phi_{01}(W_j)-\theta\} + o_p(1) \\
&\xrightarrow{d} N\left(0_{p\times 1}, \left(\frac{1+\kappa}{\kappa}\right)\left(\Sigma_{\phi(V)} + \kappa\Sigma_{\phi(W)}\right)\right).
\end{aligned}$$

The lemma is finally demonstrated by recognizing that

$$T_N = \sqrt{\frac{N}{n}}\sqrt{n}\,(U-\theta) = \sqrt{\frac{1+\kappa}{\kappa}}\sqrt{n}\,(U-\theta) + o_p(1). \quad \blacksquare$$

A.3.3 Proof of Lemma A4

R_3^* can be interpreted as a two-sample U-statistic with the kernel function

$$\phi\left(X_{1i}^A, X_{3i}^A; X_{2j}^B, X_{3j}^B\right) := X_{g,i}^A \sum_{\ell=1}^{d_2} \frac{\eta_{2\ell,j}^B K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\beta_{2\ell}}{h_\ell p_\ell\left(Z_{\ell,i}^A\right)}.$$

Subsequent derivations follow Lemma A3.

First,

$$\begin{aligned}
\phi_{10}\left(X_{1i}^A, X_{3i}^A\right) &:= E\left\{\phi\left(X_{1i}^A, X_{3i}^A; X_{2j}^B, X_{3j}^B\right)\middle| X_{1i}^A, X_{3i}^A\right\} \\
&= X_{g,i}^A \sum_{\ell=1}^{d_2} \frac{E\left\{\eta_{2\ell,j}^B K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\middle| X_{1i}^A, X_{3i}^A\right\}\beta_{2\ell}}{h_\ell p_\ell\left(Z_{\ell,i}^A\right)}.
\end{aligned}$$

In addition,

$$\begin{aligned}
&E\left\{\eta_{2\ell,j}^B K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\middle| X_{1i}^A, X_{3i}^A\right\} \\
&= E\left\{E\left(\eta_{2\ell,j}^B\middle| X_{3j}^B, X_{1i}^A, X_{3i}^A\right) K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)\middle| X_{1i}^A, X_{3i}^A\right\},
\end{aligned}$$

where by $X_{3j}^B \perp\!\!\!\perp \left(X_{1i}^A, X_{3i}^A\right)$ and Assumption S2, $E\left(\eta_{2\ell,j}^B \middle| X_{3j}^B, X_{1i}^A, X_{3i}^A\right) = E\left(\eta_{2\ell,j}^B \middle| X_{3j}^B\right) = E\left(\eta_{2\ell,j}^B \middle| Z_{\ell,j}^B\right) = 0$. Therefore, $\phi_{10}\left(X_{1i}^A, X_{3i}^A\right) \equiv 0$, and thus $\theta = E\left\{\phi_{10}\left(X_{1i}^A, X_{3i}^A\right)\right\} = 0$ also holds.

Then, defining $\phi_{01}\left(X_{2j}^B, X_{3j}^B\right) := E\left\{\phi\left(X_{1i}^A, X_{3i}^A; X_{2j}^B, X_{3j}^B\right) \middle| X_{2j}^B, X_{3j}^B\right\}$ can simplify $\sqrt{n}R_3$ as

$$\sqrt{n}R_3 = \sqrt{n}R_3^* + o_p(1) = \frac{\sqrt{\kappa}}{\sqrt{m}} \sum_{j=1}^m \phi_{01}\left(X_{2j}^B, X_{3j}^B\right) + o_p(1).$$

Therefore, deriving ψ_{21j} boils down to finding the leading term of $\phi_{01}\left(X_{2j}^B, X_{3j}^B\right)$. Now,

$$\phi_{01}\left(X_{2j}^B, X_{3j}^B\right) = \sum_{\ell=1}^{d_2} \eta_{2\ell,j}^B E\left\{ X_{g,i}^A \frac{K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)}{h_\ell p_\ell\left(Z_{\ell,i}^A\right)} \middle| X_{2j}^B, X_{3j}^B \right\} \beta_{2\ell},$$

where

$$\begin{aligned} & E\left\{ X_{g,i}^A \frac{K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)}{h_\ell p_\ell\left(Z_{\ell,i}^A\right)} \middle| X_{2j}^B, X_{3j}^B \right\} \\ &= E\left\{ E\left(X_{g,i}^A \middle| Z_{\ell,i}^A, X_{2j}^B, X_{3j}^B\right) \frac{K\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right)}{h_\ell p_\ell\left(Z_{\ell,i}^A\right)} \middle| X_{2j}^B, X_{3j}^B \right\} \qquad \text{(A7)} \end{aligned}$$

and $E\left(X_{g,i}^A \middle| Z_{\ell,i}^A, X_{2j}^B, X_{3j}^B\right) = E\left(X_{g,i}^A \middle| Z_{\ell,i}^A\right)$ by $Z_{\ell,i}^A \perp\!\!\!\perp \left(X_{2j}^B, X_{3j}^B\right)$. Then, by symmetry of $K(\cdot)$, the definition of $G_\ell^A(\cdot)$, a change of variable $\left(t - Z_{\ell,i}^A\right)/h_\ell =: u$, and Assumption P4, the right-hand side of (A7) collapses to

$$\frac{1}{h_\ell} \int G_\ell^A(t) \frac{K\left(\frac{t - Z_{\ell,j}^B}{h_\ell}\right)}{p_\ell(t)} p_\ell(t)\, dt = \int G_\ell^A\left(Z_{\ell,j}^B + h_\ell u\right) K(u)\, du = G_\ell^A\left(Z_{\ell,j}^B\right) + O\left(h^\nu\right).$$

It follows that $\phi_{01}\left(X_{2j}^B, X_{3j}^B\right) = \sum_{\ell=1}^{d_2} G_\ell^A\left(Z_{\ell,j}^B\right) \eta_{2\ell,j}^B \beta_{2\ell} + O_p\left(h^\nu\right)$ uniformly on $\mathbb{Z}$, and thus

$$\frac{\sqrt{\kappa}}{\sqrt{m}} \sum_{j=1}^m \phi_{01}\left(X_{2j}^B, X_{3j}^B\right) = \frac{\sqrt{\kappa}}{\sqrt{m}} \sum_{j=1}^m \sum_{\ell=1}^{d_2} G_\ell^A\left(Z_{\ell,j}^B\right) \eta_{2\ell,j}^B \beta_{2\ell} + O_p\left(\sqrt{m} h^\nu\right),$$

where $\sqrt{m} h^\nu = O\left(\sqrt{n h^{2\nu}}\right) = o(1)$ by $m \asymp n$. This completes the proof. ■

A.3.4 Proof of Lemma A5

A mean-value expansion of $K\left\{\left(\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}\right)/h_{\ell}\right\}$ around $\left(\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}\right)/h_{\ell}=\left(Z_{\ell,j}^{B}-Z_{\ell,i}^{A}\right)/h_{\ell}$ yields

$$
\begin{aligned}
&K\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-K\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\\
&=K^{(1)}\left(\overline{\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}}\right)\left\{\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right\}\\
&=K^{(1)}\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\left\{\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right\}\\
&+\left\{K^{(1)}\left(\overline{\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}}\right)-K^{(1)}\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right\}\\
&\times\left\{\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right\} \qquad \text{(A8)}
\end{aligned}
$$

for some $\left(\overline{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}\right)/h_{\ell}$ on the line segment joining $\left(\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}\right)/h_{\ell}$ and $\left(Z_{\ell,j}^{B}-Z_{\ell,i}^{A}\right)/h_{\ell}$. By the definition of indices,

$$
\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)=\frac{1}{h_{\ell}}\left(X_{3j}^{B}-X_{3i}^{A}\right)^{\top}\left(\hat{\theta}_{\ell}-\theta_{\ell}\right),
$$

so that

$$
\begin{aligned}
&K^{(1)}\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\left\{\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right\}\\
&=K^{(1)}\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\frac{1}{h_{\ell}}\left(X_{3j}^{B}-X_{3i}^{A}\right)^{\top}\left(\hat{\theta}_{\ell}-\theta_{\ell}\right). \qquad \text{(A9)}
\end{aligned}
$$

Moreover, by Assumptions S1, S7 and P3,

$$
\begin{aligned}
\left|\left(\frac{\hat{Z}_{\ell,j}^{B}-\hat{Z}_{\ell,i}^{A}}{h_{\ell}}\right)-\left(\frac{Z_{\ell,j}^{B}-Z_{\ell,i}^{A}}{h_{\ell}}\right)\right| &\leq \frac{1}{h_{\ell}}\times 2\sup_{x_3\in\mathbb{X}_{3C}\times\mathbb{X}_{3D}}\|x_3\|\left\|\hat{\theta}_{\ell}-\theta_{\ell}\right\|\\
&=O_p\left(\frac{1}{\sqrt{mh^2}}\right).
\end{aligned}
$$

It also follows from Assumption P2 that

$$\begin{aligned}
\left|K^{(1)}\left(\overline{\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}}\right)-K^{(1)}\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right| &\le L\left|\left(\overline{\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}}\right)-\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right| \\
&\le L\left|\left(\frac{\hat{Z}^B_{\ell,j}-\hat{Z}^A_{\ell,i}}{h_\ell}\right)-\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right| \\
&= O_p\left(\frac{1}{\sqrt{mh^2}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left\{K^{(1)}\left(\overline{\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}}\right)-K^{(1)}\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right\} \\
&\times\left\{\left(\frac{\hat{Z}^B_{\ell,j}-\hat{Z}^A_{\ell,i}}{h_\ell}\right)-\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right\} \\
&= O_p\left(\frac{1}{mh^2}\right) \qquad \text{(A10)}
\end{aligned}$$

uniformly on $\mathbb{Z}$. Substituting (A9) and (A10) into (A8) establishes the lemma. ■

A.3.5 Proof of Lemma A6

A mean-value expansion of $\hat{g}_{2\ell}\left(\hat{Z}^A_{\ell,i}\right)=\hat{r}_\ell\left(\hat{Z}^A_{\ell,i}\right)/\hat{p}_\ell\left(\hat{Z}^A_{\ell,i}\right)$ around $\hat{g}_{2\ell}\left(\hat{Z}^A_{\ell,i}\right)=\tilde{g}_{2\ell}\left(Z^A_{\ell,i}\right)=\tilde{r}_\ell\left(Z^A_{\ell,i}\right)/\tilde{p}_\ell\left(Z^A_{\ell,i}\right)$ yields

$$\begin{aligned}
\hat{g}_{2\ell}\left(\hat{Z}^A_{\ell,i}\right)-\tilde{g}_{2\ell}\left(Z^A_{\ell,i}\right) &= \left\{\frac{1}{\overline{p_\ell\left(Z^A_{\ell,i}\right)}}\right\}\left\{\hat{r}_\ell\left(\hat{Z}^A_{\ell,i}\right)-\tilde{r}_\ell\left(Z^A_{\ell,i}\right)\right\} \\
&\quad -\left\{\frac{\overline{r_\ell\left(Z^A_{\ell,i}\right)}}{\overline{p_\ell\left(Z^A_{\ell,i}\right)}^2}\right\}\left\{\hat{p}_\ell\left(\hat{Z}^A_{\ell,i}\right)-\tilde{p}_\ell\left(Z^A_{\ell,i}\right)\right\} \qquad \text{(A11)}
\end{aligned}$$

for some $\overline{q_\ell\left(Z^A_{\ell,i}\right)}$ on the line segment joining $\hat{q}_\ell\left(\hat{Z}^A_{\ell,i}\right)$ and $\tilde{q}_\ell\left(Z^A_{\ell,i}\right)$ for $q_\ell\in\{r_\ell,p_\ell\}$. Applying Lemma A5 to

$$\hat{r}_\ell\left(\hat{Z}^A_{\ell,i}\right)-\tilde{r}_\ell\left(Z^A_{\ell,i}\right)=\frac{1}{mh_\ell}\sum_{j=1}^m X^B_{2\ell,j}\left\{K\left(\frac{\hat{Z}^B_{\ell,j}-\hat{Z}^A_{\ell,i}}{h_\ell}\right)-K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right\}, \text{ and}$$

$$\hat{p}_\ell\left(\hat{Z}^A_{\ell,i}\right)-\tilde{p}_\ell\left(Z^A_{\ell,i}\right)=\frac{1}{mh_\ell}\sum_{j=1}^m \left\{K\left(\frac{\hat{Z}^B_{\ell,j}-\hat{Z}^A_{\ell,i}}{h_\ell}\right)-K\left(\frac{Z^B_{\ell,j}-Z^A_{\ell,i}}{h_\ell}\right)\right\},$$

we have

$$\hat{r}_\ell\left(\hat{Z}_{\ell,i}^A\right)-\tilde{r}_\ell\left(Z_{\ell,i}^A\right):=\Lambda_{\ell,i}\left(\hat{\theta}_\ell-\theta_\ell\right)+O_p\left(\frac{1}{mh^3}\right),\ \text{and} \tag{A12}$$

$$\hat{p}_\ell\left(\hat{Z}_{\ell,i}^A\right)-\tilde{p}_\ell\left(Z_{\ell,i}^A\right):=\Delta_{\ell,i}\left(\hat{\theta}_\ell-\theta_\ell\right)+O_p\left(\frac{1}{mh^3}\right), \tag{A13}$$

where

$$\Lambda_{\ell,i}:=\frac{1}{mh_\ell^2}\sum_{j=1}^{m}X_{2\ell,j}^B K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)\left(X_{3j}^B-X_{3i}^A\right)^\top, \tag{A14}$$

$$\Delta_{\ell,i}:=\frac{1}{mh_\ell^2}\sum_{j=1}^{m}K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)\left(X_{3j}^B-X_{3i}^A\right)^\top, \tag{A15}$$

and it will be demonstrated shortly that each of $\Lambda_{\ell,i}$ and $\Delta_{\ell,i}$ is $O_p(1)$ uniformly on $\mathbb{Z}_\ell$. As a result, each of $\left|\overline{r_\ell\left(Z_{\ell,i}^A\right)}-\tilde{r}_\ell\left(Z_{\ell,i}^A\right)\right|$ and $\left|\overline{p_\ell\left(Z_{\ell,i}^A\right)}-\tilde{p}_\ell\left(Z_{\ell,i}^A\right)\right|$ is bounded by $O_p\left(m^{-1/2}\right)+O_p\left\{1/\left(mh^3\right)\right\}=O_p\left(m^{-1/2}\right)$ uniformly on $\mathbb{Z}_\ell$ by $1/\left(mh^6\right)\to 0$. Therefore, by the definition of $\tilde{r}_\ell\left(Z_{\ell,i}^A\right)$,

$$\frac{1}{\overline{p_\ell\left(Z_{\ell,i}^A\right)}}=\frac{1}{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)}+O_p\left(\frac{1}{\sqrt{m}}\right),\ \text{and} \tag{A16}$$

$$\frac{\overline{r_\ell\left(Z_{\ell,i}^A\right)}}{\overline{p_\ell\left(Z_{\ell,i}^A\right)}^2}=\frac{\tilde{r}_\ell\left(Z_{\ell,i}^A\right)}{\left\{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)\right\}^2}+O_p\left(\frac{1}{\sqrt{m}}\right)=\frac{\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)}{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)}+O_p\left(\frac{1}{\sqrt{m}}\right). \tag{A17}$$

Substituting (A12)-(A17) into (A11), we finally obtain

$$\begin{aligned}&\hat{g}_{2\ell}\left(\hat{Z}_{\ell,i}^A\right)-\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)\\&=\left\{\frac{1}{\tilde{p}_\ell\left(Z_{\ell,i}^A\right)}\right\}\left\{\Lambda_{\ell,i}-\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)\Delta_{\ell,i}\right\}\left(\hat{\theta}_\ell-\theta_\ell\right)+O_p\left(\frac{1}{mh^3}\right).\end{aligned} \tag{A18}$$

The remaining task is to approximate $\Lambda_{\ell,i}$ and $\Delta_{\ell,i}$. For the latter, consider the identity $\Delta_{\ell,i}\equiv E\left(\Delta_{\ell,i}\mid X_{3i}^A\right)+\left\{\Delta_{\ell,i}-E\left(\Delta_{\ell,i}\mid X_{3i}^A\right)\right\}$. In particular,

$$\begin{aligned}E\left(\Delta_{\ell,i}\mid X_{3i}^A\right)&=\frac{1}{h_\ell^2}E\left\{K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)\left(X_{3j}^B-X_{3i}^A\right)^\top\middle|X_{3i}^A\right\}\\&=\frac{1}{h_\ell^2}E\left[K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)\left\{E\left(X_{3j}^B\mid Z_{\ell,j}^B,X_{3i}^A\right)-X_{3i}^A\right\}^\top\middle|X_{3i}^A\right]\\&=\frac{1}{h_\ell^2}E\left\{K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)g_{3\ell}^B\left(Z_{\ell,j}^B\right)^\top\middle|X_{3i}^A\right\}\\&\quad-\frac{1}{h_\ell^2}E\left\{K^{(1)}\left(\frac{Z_{\ell,j}^B-Z_{\ell,i}^A}{h_\ell}\right)\middle|X_{3i}^A\right\}X_{3i}^{A\top}\end{aligned}$$

by $Z_{\ell,j}^B \perp\!\!\!\perp X_{3i}^A$ so that $E\left(X_{3j}^B\middle| Z_{\ell,j}^B, X_{3i}^A\right) = E\left(X_{3j}^B\middle| Z_{\ell,j}^B\right) = g_{3\ell}^B\left(Z_{\ell,j}^B\right)$. Moreover,

$$\begin{aligned}
&\frac{1}{h_\ell^2} E\left\{ K^{(1)}\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right) g_{3\ell}^B\left(Z_{\ell,j}^B\right)^\top \middle| X_{3i}^A \right\} \\
&= \frac{1}{h_\ell^2}\int K^{(1)}\left(\frac{t - Z_{\ell,i}^A}{h_\ell}\right) g_{3\ell}^B(t)^\top p_\ell(t)\,dt \\
&= \frac{1}{h_\ell}\int K^{(1)}(u)\, g_{3\ell}^B\left(Z_{\ell,i}^A + h_\ell u\right)^\top p_\ell\left(Z_{\ell,i}^A + h_\ell u\right) du \\
&= -\int K(u)\left\{ g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A + h_\ell u\right)^\top p_\ell\left(Z_{\ell,i}^A + h_\ell u\right) + g_{3\ell}^B\left(Z_{\ell,i}^A + h_\ell u\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A + h_\ell u\right)\right\} du \\
&= -\left\{ g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right) + g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} + O\left(h^\nu\right)
\end{aligned}$$

by a change of variable $\left(t - Z_{\ell,i}^A\right)/h_\ell =: u$ and integration by parts. Similarly,

$$\frac{1}{h_\ell^2} E\left\{ K^{(1)}\left(\frac{Z_{\ell,j}^B - Z_{\ell,i}^A}{h_\ell}\right) \middle| X_{3i}^A \right\} = -p_\ell^{(1)}\left(Z_{\ell,i}^A\right) + O\left(h^\nu\right),$$

and thus

$$\begin{aligned}
&E\left(\Delta_{\ell,i}\middle| X_{3i}^A\right) \\
&= -\left\{ g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right) + g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} + X_{3i}^{A\top} p_\ell^{(1)}\left(Z_{\ell,i}^A\right) + O\left(h^\nu\right).
\end{aligned}$$

Because Lemma A1 also implies that $\Delta_{\ell,i} - E\left(\Delta_{\ell,i}\middle| X_{3i}^A\right) = O_p\left(\sqrt{\log m/\left(mh^3\right)}\right)$ uniformly on $\mathbb{Z}_\ell$, we finally have

$$\begin{aligned}
\Delta_{\ell,i} = &-\left\{ g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right) + g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} \\
&+ X_{3i}^{A\top} p_\ell^{(1)}\left(Z_{\ell,i}^A\right) + O_p\left(h^\nu + \sqrt{\frac{\log m}{mh^3}}\right) \\
&= -\left\{ g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right) + g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} \\
&+ X_{3i}^{A\top} p_\ell^{(1)}\left(Z_{\ell,i}^A\right) + o_p(1). \qquad \text{(A19)}
\end{aligned}$$

A similar argument also establishes that

$$\begin{aligned}
\Lambda_{\ell,i} = &-\left\{ g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right) g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right) + g_{2\ell}\left(Z_{\ell,i}^A\right) g_{3\ell}^{B(1)}\left(Z_{\ell,i}^A\right)^\top p_\ell\left(Z_{\ell,i}^A\right)\right. \\
&\left. + g_{2\ell}\left(Z_{\ell,i}^A\right) g_{3\ell}^B\left(Z_{\ell,i}^A\right)^\top p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} \\
&+ \left\{ g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right) p_\ell\left(Z_{\ell,i}^A\right) + g_{2\ell}\left(Z_{\ell,i}^A\right) p_\ell^{(1)}\left(Z_{\ell,i}^A\right)\right\} X_{3i}^{A\top} + o_p(1). \qquad \text{(A20)}
\end{aligned}$$

Substituting (A19) and (A20) into (A18) and using $\tilde{p}_\ell\left(Z_{\ell,i}^A\right) = p_\ell\left(Z_{\ell,i}^A\right) + o_p(1)$ and $\tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right) = g_{2\ell}\left(Z_{\ell,i}^A\right) + o_p(1)$ uniformly on $\mathbb{Z}_\ell$ by Lemma A1, we obtain

$$\begin{aligned}
&\hat{g}_{2\ell}\left(\hat{Z}_{\ell,i}^A\right) - \tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right) \\
&= \left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\left(\hat{\theta}_\ell - \theta_\ell\right) + o_p\left(\frac{1}{\sqrt{m}}\right) + O_p\left(\frac{1}{mh^3}\right),
\end{aligned}$$

where $1/\left(mh^3\right) = o\left(m^{-1/2}\right)$ because $1/\left(mh^6\right) = o(1)$. Then, by $n/m = \kappa + o(1)$, it can be concluded that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{g}_{2\ell}\left(\hat{Z}_{\ell,i}^A\right) - \tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)\right\} \\
&= \sqrt{\frac{n}{m}}\left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\sqrt{m}\left(\hat{\theta}_\ell - \theta_\ell\right) + o_p(1) \\
&= \sqrt{\kappa}\left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\sqrt{m}\left(\hat{\theta}_\ell - \theta_\ell\right) + o_p(1). \quad \blacksquare
\end{aligned}$$

A.3.6 Proof of Lemma A7

It follows from the definition of R_4 that

$$\begin{aligned}
\sqrt{n}R_4 &= \frac{1}{n}\sum_{i=1}^n \hat{X}_{g,i}^A\sqrt{n}\left\{\hat{g}_2\left(\hat{Z}_i^A\right) - \tilde{g}_2\left(Z_i^A\right)\right\}^\top \beta_2 \\
&= \sum_{\ell=1}^{d_2}\frac{1}{n}\sum_{i=1}^n X_{g,i}^A\sqrt{n}\left\{\hat{g}_{2\ell}\left(\hat{Z}_{\ell,i}^A\right) - \tilde{g}_{2\ell}\left(Z_{\ell,i}^A\right)\right\}\beta_{2\ell} + o_p(1),
\end{aligned}$$

where the $o_p(1)$ rate is uniform on $\mathbb{Z}$. Substituting Lemma A6 into the right-hand side and then using Eq. (11) of the main text, we have

$$\begin{aligned}
\sqrt{n}A_4 &= \sum_{\ell=1}^{d_2}\frac{1}{n}\sum_{i=1}^n X_{g,i}^A\left[\sqrt{\kappa}\left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\sqrt{m}\left(\hat{\theta}_\ell - \theta_\ell\right)\right]\beta_{2\ell} + o_p(1) \\
&= \sqrt{\kappa}\sum_{\ell=1}^{d_2}\left[\frac{1}{n}\sum_{i=1}^n X_{g,i}^A\left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right)\right]\frac{1}{\sqrt{m}}\sum_{j=1}^m \varphi_{\ell j}^B(\theta_\ell)\beta_{2\ell} + o_p(1).
\end{aligned}$$

The result immediately follows from

$$\frac{1}{n}\sum_{i=1}^n X_{g,i}^A\left\{X_{3i}^A - g_{3\ell}^B\left(Z_{\ell,i}^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_{\ell,i}^A\right) = E\left[X_g^A\left\{X_3^A - g_{3\ell}^B\left(Z_\ell^A\right)\right\}^\top g_{2\ell}^{(1)}\left(Z_\ell^A\right)\right] + o_p(1). \quad \blacksquare$$

A.3.7 Proof of Theorem 2

It follows from (A4) and Assumption P3 that

$$\sqrt{n}\left(\hat{\beta}_{PS} - \beta\right) = Q_{\hat{X}_g^A\hat{X}_g^A}^{-1}\left\{\sqrt{n}R_2 - \sqrt{n}\left(R_3 + R_4\right)\right\} + o_p(1).$$

Now, $Q_{\hat{X}_g^A \hat{X}_g^A} \xrightarrow{p} \Phi_{PS} = E\left(X_g^A X_g^{A\top}\right)$, and Φ_{PS} is positive definite by Assumption P1. By CLT,

$$\sqrt{n}R_2 = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_{1i} + o_p(1) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} X_{gi}^A \epsilon_i^A + o_p(1) \xrightarrow{d} N\left(0_{(d+1)\times 1}, \Omega_{PS,1}\right).$$

By Lemmata A4 and A7,

$$\sqrt{n}\left(R_3 + R_4\right) = \frac{\sqrt{\kappa}}{\sqrt{m}}\sum_{j=1}^{m}\psi_{2j} + o_p(1) \xrightarrow{d} N\left(0_{(d+1)\times 1}, \kappa\Omega_{PS,2}\right).$$

The theorem is finally established by independence between two influence functions ψ_{1i} and ψ_{2j}. ■

Supplement B: Covariance Estimation

Estimating the asymptotic covariance matrix V_{PS} in Theorem 2 is essential for inference. Since $Q_{\hat{X}_g^A \hat{X}_g^A}$ is consistent for Φ_{PS} and κ may be replaced by n/m, the problem of estimating V_{PS} consistently boils down to proposing consistent estimators of two covariance matrices $\Omega_{PS,1}$ and $\Omega_{PS,2}$. Let $\hat{\Omega}_{PS,1}$ and $\hat{\Omega}_{PS,2}$ denote their consistent estimates. Then, V_{PS} can be estimated by

$$\hat{V}_{PS} := Q^{-1}_{\hat{X}_g^A \hat{X}_g^A}\hat{\Omega}_{PS} Q^{-1}_{\hat{X}_g^A \hat{X}_g^A} := Q^{-1}_{\hat{X}_g^A \hat{X}_g^A}\left(\hat{\Omega}_{PS,1} + \frac{n}{m}\hat{\Omega}_{PS,2}\right) Q^{-1}_{\hat{X}_g^A \hat{X}_g^A}.$$

Consistency of $\hat{V}_{PS}$ can be understood in line with the proofs of Theorems 1 and 2 in Supplement A. In what follows, we explain how to obtain $\hat{\Omega}_{PS,1}$ and $\hat{\Omega}_{PS,2}$ separately. It turns out that the resulting covariance estimate $\hat{V}_{PS}$ is a hybrid of analytical and resampling-based methods.

B.1 Estimation of $\Omega_{PS,1}$

It is straightforward to estimate $\Omega_{PS,1}$. Given the PILS-SDR residual $\hat{\epsilon}_i^A := Y_i^A - \hat{X}_{g,i}^{A\top}\hat{\beta}_{PS}$, we can estimate $\Omega_{PS,1}$ consistently by

$$\hat{\Omega}_{PS,1} := \frac{1}{n}\sum_{i=1}^{n}\hat{X}_{g,i}^A \hat{X}_{g,i}^{A\top}\left(\hat{\epsilon}_i^A\right)^2.$$

B.2 Estimation of $\Omega_{PS,2}$

In contrast, consistent estimation of $\Omega_{PS,2}$ is highly complicated. Indeed, estimating each of two influence functions ψ_{21j} and ψ_{22j} takes multiple steps. Let $\hat{\psi}_{21j}$ and $\hat{\psi}_{22j}$

be estimates of ψ_{21j} and ψ_{22j}, respectively, so that $\psi_{2j} = \psi_{21j} + \psi_{22j}$ can be estimated by $\hat{\psi}_{2j} := \hat{\psi}_{21j} + \hat{\psi}_{22j}$. This leads to a consistent estimator of $\Omega_{PS,2}$ as

$$\hat{\Omega}_{PS,2} := \frac{1}{m} \sum_{j=1}^{m} \hat{\psi}_{2j} \hat{\psi}_{2j}^{\top}.$$

In what follows, we describe detailed procedures of computing $\hat{\psi}_{21j}$ and $\hat{\psi}_{22j}$.

B.2.1 Computation Details of $\hat{\psi}_{21j}$

A general form of $\hat{\psi}_{21j}$ is given by

$$\hat{\psi}_{21j} := \sum_{\ell=1}^{d_2} \hat{G}_{\ell}^{A} \left(\hat{Z}_{\ell,j}^{B} \right) \hat{\eta}_{2\ell,j}^{B} \hat{\beta}_{PS,2,\ell},$$

where $\hat{G}_{\ell}^{A} \left(\hat{Z}_{\ell,j}^{B} \right)$ is an estimate of $G_{\ell}^{A} \left(Z_{\ell,j}^{B} \right)$ to be discussed shortly, and $\hat{\beta}_{PS,2} = \left(\hat{\beta}_{PS,2,1}, \ldots, \hat{\beta}_{PS,2,d_2} \right)^{\top}$ is the PILS-SDR estimate of β_2. The nonparametric regression residual $\hat{\eta}_{2\ell,j}^{B}$ can be computed as $\hat{\eta}_{2\ell,j}^{B} := X_{2\ell,j}^{B} - \hat{g}_{2\ell,-j}^{B} \left(\hat{Z}_{\ell,j}^{B} \right)$, where

$$\hat{g}_{2\ell,-j}^{B} (z) := \frac{\sum_{k=1,k\neq j}^{m} X_{2\ell,k}^{B} K \left(\frac{\hat{Z}_{\ell,k}^{B} - z}{\hat{h}_{\ell}} \right)}{\sum_{k=1,k\neq j}^{m} K \left(\frac{\hat{Z}_{\ell,k}^{B} - z}{\hat{h}_{\ell}} \right)}$$

is the leave-one-out Nadaraya-Watson (NW) estimator with the j-th observation dropped, and $\hat{h}_{\ell}$ is the rule-of-thumb bandwidth from Eq. (8) of the main text. The reason for employing the leave-one-out NW estimator is to avoid over-fitting the predicted value of $X_{2\ell,j}^{B}$; if, otherwise, the same observations on $X_{2\ell,j}^{B}$ were also used to train the kernel, an over-fitted NW estimate of $g_{2\ell}^{B} \left(Z_{\ell,j}^{B} \right)$ would lead to an under-fitted residual and then an under-estimate of $\Omega_{PS,2}$.

Now, $G_{\ell}^{A} \left(Z_{\ell,j}^{B} \right)$ for a fixed $Z_{\ell,j}^{B}$ can be expressed precisely as

$$\begin{aligned} & G_{\ell}^{A} \left(Z_{\ell,j}^{B} \right) \\ & = \left(1, E \left(X_1^A \middle| Z_{\ell}^{A} = Z_{\ell,j}^{B} \right)^{\top}, E \left\{ g_2 \left(Z^A \right) \middle| Z_{\ell}^{A} = Z_{\ell,j}^{B} \right\}^{\top}, E \left(X_{3I}^{A} \middle| Z_{\ell}^{A} = Z_{\ell,j}^{B} \right)^{\top} \right)^{\top}. \end{aligned}$$

Therefore, $\hat{G}_{\ell}^{A} \left(\hat{Z}_{\ell,j}^{B} \right)$ is in the form of

$$\begin{aligned} & \hat{G}_{\ell}^{A} \left(\hat{Z}_{\ell,j}^{B} \right) \\ & := \left(1, \hat{E} \left(X_1^A \middle| \hat{Z}_{\ell}^{A} = \hat{Z}_{\ell,j}^{B} \right)^{\top}, \hat{E} \left\{ \hat{g}_2 \left(\hat{Z}^A \right) \middle| \hat{Z}_{\ell}^{A} = \hat{Z}_{\ell,j}^{B} \right\}^{\top}, \hat{E} \left(X_{3I}^{A} \middle| \hat{Z}_{\ell}^{A} = \hat{Z}_{\ell,j}^{B} \right)^{\top} \right)^{\top}. \end{aligned}$$

Below we deliver definitions of $\hat{E}\left(X_1^A \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right)$, $\hat{E}\left\{\hat{g}_2\left(\hat{Z}^A\right) \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right\}$ and $\hat{E}\left(X_{3I}^A \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right)$.

First, using $\left\{\left(X_{1i}^A, X_{3I,i}^A, \hat{Z}_i^A\right)\right\}_{i=1}^n$ and $\left\{\hat{Z}_j^B\right\}_{j=1}^m$ as data and design points, respectively, we define $\hat{E}\left(X_1^A \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right)$ as a NW regression estimate

$$\hat{E}\left(X_1^A \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right) := \frac{\frac{1}{nb_\ell}\sum_{i=1}^n X_{1i}^A K\left(\frac{\hat{Z}_{\ell,i}^A - \hat{Z}_{\ell,j}^B}{b_\ell}\right)}{\frac{1}{nb_\ell}\sum_{i=1}^n K\left(\frac{\hat{Z}_{\ell,i}^A - \hat{Z}_{\ell,j}^B}{b_\ell}\right)}$$

In practice, the twiced Gaussian kernel in Eq. (7) of the main text and the rule-of-thumb bandwidth

$$\hat{b}_\ell := \frac{1}{2}\hat{\sigma}_{\hat{Z}_\ell^A}\left(\frac{\log n}{n}\right)^{\bar{\alpha}}$$

can be employed for $K(u)$ and b_ℓ, where $\hat{\sigma}_{\hat{Z}_\ell^A}$ is the sample standard deviation of $\left\{\hat{Z}_{\ell,i}^A\right\}_{i=1}^n$ and $\bar{\alpha} = 3/20$.

Second, to estimate $E\left\{g_2\left(Z^A\right) \middle| Z_\ell^A = Z_{\ell,j}^B\right\}$, we should consider two cases. Observe that $E\left\{g_{2k}\left(Z_k^A\right) \middle| Z_\ell^A = Z_{\ell,j}^B\right\}$ collapses to $g_{2\ell}\left(Z_{\ell,j}^B\right)$ for $k = \ell$. In this case, the conditional expectation may be simply replaced by $\hat{g}_{2\ell}\left(\hat{Z}_{\ell,j}^B\right)$. On the other hand, $E\left\{g_{2k}\left(Z_k^A\right) \middle| Z_\ell^A = Z_{\ell,j}^B\right\}$ for $k \neq \ell$ can be again estimated by NW regression smoothing as

$$\hat{E}\left\{\hat{g}_{2k}\left(\hat{Z}_k^A\right) \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right\} := \frac{\frac{1}{nb_\ell}\sum_{i=1}^n \hat{g}_{2k}\left(\hat{Z}_{k,i}^A\right) K\left(\frac{\hat{Z}_{\ell,i}^A - \hat{Z}_{\ell,j}^B}{b_\ell}\right)}{\frac{1}{nb_\ell}\sum_{i=1}^n K\left(\frac{\hat{Z}_{\ell,i}^A - \hat{Z}_{\ell,j}^B}{b_\ell}\right)}.$$

The same kernel and bandwidth as above apply for implementation. It follows that $E\left\{g_2\left(Z^A\right) \middle| Z_\ell^A = Z_{\ell,j}^B\right\}$ can be estimated by

$$\begin{aligned}&\hat{E}\left\{\hat{g}_2\left(\hat{Z}^A\right) \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right\}\\ &:= \left(\hat{E}\left\{\hat{g}_{21}\left(\hat{Z}_1^A\right) \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right\}, \ldots, \hat{g}_{2\ell}\left(\hat{Z}_{\ell,j}^B\right), \ldots, \hat{E}\left\{\hat{g}_{2d_2}\left(\hat{Z}_{d_2}^A\right) \middle| \hat{Z}_\ell^A = \hat{Z}_{\ell,j}^B\right\}\right)^\top.\end{aligned}$$

Third, observe that $E\left(X_{3I}^A \middle| Z_\ell^A = Z_{\ell,j}^B\right)$ is a subset of $g_{3\ell}^A\left(Z_{\ell,j}^B\right) = E\left(X_3^A \middle| Z_\ell^A = Z_{\ell,j}^B\right)$. Invoke that X_3^A is linear in Z_ℓ^A under the condition (ii) of Assumption S2. This motivates us to express $g_{3\ell}^A\left(Z_\ell^A\right) = \delta_{0\ell}^A + \delta_{1\ell}^A Z_\ell^A$ for some $d_3 \times 1$ slope and intercept vectors $\delta_{0\ell}^A$ and $\delta_{1\ell}^A$, respectively. These vectors can be estimated by OLS for the regression of X_3^A on $\left(1, \hat{Z}_\ell^A\right)$ in the same equation-by-equation manner as in seemingly unrelated regressions. Let OLS estimates of $\delta_{0\ell}^{A}$ and $\delta_{1\ell}^{A}$ be denoted by $\hat{\delta}_{0\ell}^{A}$ and $\hat{\delta}_{1\ell}^{A}$, respectively. Then, $\hat{E}\left(X_{3I}^{A}\middle|\hat{Z}_{\ell}^{A}=\hat{Z}_{\ell,j}^{B}\right)$ can be obtained as the corresponding subset of $\hat{g}_{3\ell}^{A}\left(\hat{Z}_{\ell,j}^{B}\right)=\hat{\delta}_{0\ell}^{A}+\hat{\delta}_{1\ell}^{A}\hat{Z}_{\ell,j}^{B}$, i.e., OLS fitted values with the predictor $\hat{Z}_{\ell,j}^{B}$ plugged in.

B.2.2 Computation Details of $\hat{\psi}_{22j}$

Observe that $\sum_{i=1}^{n}\hat{X}_{g,i}^{A}\left\{X_{3i}^{A}-\hat{g}_{3\ell}^{B}\left(\hat{Z}_{\ell,i}^{A}\right)\right\}^{\top}\hat{g}_{2\ell}^{(1)}\left(\hat{Z}_{\ell,i}^{A}\right)/n$ is the sample counterpart of $E\left[X_{g}^{A}\left\{X_{3}^{A}-g_{3\ell}^{B}\left(Z_{\ell}^{A}\right)\right\}^{\top}g_{2\ell}^{(1)}\left(Z_{\ell}^{A}\right)\right]$, where the definitions of $\hat{g}_{3\ell}^{B}\left(\hat{Z}_{\ell,i}^{A}\right)$ and $\hat{g}_{2\ell}^{(1)}\left(\hat{Z}_{\ell,i}^{A}\right)$ will be revealed shortly. Also let $\hat{\varphi}_{\ell,j}^{B}\left(\hat{\theta}_{\ell}\right)$ be an estimate of the influence function $\varphi_{\ell,j}^{B}\left(\theta_{\ell}\right)$, where specific forms of $\hat{\varphi}_{\ell,j}^{B}\left(\hat{\theta}_{\ell}\right)$ will be delivered later. Given these notations, $\hat{\psi}_{22j}$ can be expressed as

$$\hat{\psi}_{22j}:=\sum_{\ell=1}^{d_2}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{X}_{g,i}^{A}\left\{X_{3i}^{A}-\hat{g}_{3\ell}^{B}\left(\hat{Z}_{\ell,i}^{A}\right)\right\}^{\top}\hat{g}_{2\ell}^{(1)}\left(\hat{Z}_{\ell,i}^{A}\right)\right]\hat{\varphi}_{\ell,j}^{B}\left(\hat{\theta}_{\ell}\right)\hat{\beta}_{PS,2,\ell}.$$

Obtaining $\hat{g}_{3\ell}^{B}\left(\hat{Z}_{\ell,i}^{A}\right)$ largely follows the previous section. Again, under the linearity condition, we may express $g_{3\ell}^{B}\left(Z_{\ell}^{B}\right)=\delta_{0\ell}^{B}+\delta_{1\ell}^{B}Z_{\ell}^{B}$ for some $d_3\times 1$ slope and intercept vectors $\delta_{0\ell}^{B}$ and $\delta_{1\ell}^{B}$, respectively. Then, an equation-by-equation OLS for the regression of X_{3}^{B} on $\left(1,\hat{Z}_{\ell}^{B}\right)$ yields their estimates $\hat{\delta}_{0\ell}^{B}$ and $\hat{\delta}_{1\ell}^{B}$. As a consequence, $\hat{g}_{3\ell}^{B}\left(\hat{Z}_{\ell,i}^{A}\right)$ is a set of OLS fitted values with the predictor $\hat{Z}_{\ell,i}^{A}$ plugged in.

We adopt the two-sided finite-difference (FD) method by Ullah (1988) to compute the first-order regression derivative estimate $\hat{g}_{2\ell}^{(1)}\left(\hat{Z}_{\ell,i}^{A}\right)$. The FD estimator takes the form of

$$\hat{g}_{2\ell}^{(1)}(z):=\frac{1}{2h_{\ell}}\left\{\hat{g}_{2\ell}\left(z+h_{\ell}\right)-\hat{g}_{2\ell}\left(z-h_{\ell}\right)\right\}. \quad \text{(B1)}$$

Again the twiced Gaussian kernel (given in Eq. (7) of the main text) and the rule-of-thumb bandwidth (given in Eq. (8) of the main text) are employed for (B1).

The remaining task is to estimate the influence function $\varphi_{\ell,j}^{B}\left(\theta_{\ell}\right)$. Since its exact expression for SIR, PLS and PIR is highly complicated in general, we rely on a particular resampling method. More specifically, $\varphi_{\ell,j}^{B}\left(\theta_{\ell}\right)$ is replaced with the jackknife influence function $(m-1)\left(\overline{\hat{\theta}_{\ell}^{\bullet}}-\hat{\theta}_{\ell,-j}^{\bullet}\right)$ from Efron (1992), where $\hat{\theta}_{\ell,-j}^{\bullet}$ for $\bullet\in\{SIR,PLS,PIR\}$ is the *delete-one jackknife estimate* of θ_{ℓ} (e.g., Shao, 1989) obtained by eliminating the jth observation from $\mathcal{S}^{B}$ but maintaining values of user-specific selection parameters (i.e., H and/or q) at those chosen in full-sample cases, and $\overline{\hat{\theta}_{\ell}^{\bullet}}:=\sum_{j=1}^{m}\hat{\theta}_{\ell,-j}^{\bullet}/m$.

For Probit and Logit, we consider influence functions under misspecified maximum likelihood estimation (MLE) by White (1982). It is assumed that an intercept enters each index, and we estimate the amended index $\mathcal{Z}_\ell^B := \gamma_\ell^\top \mathcal{X}_3^B$ for $\gamma_\ell := \left(\theta_\ell^\top, \alpha_\ell\right)^\top$ and $\mathcal{X}_3^B := \left(X_3^{B\top}, 1\right)^\top$. Clearly, $\mathcal{Z}_\ell^B = Z_\ell^B + \alpha_\ell$ holds, and the link function estimation in Section 2.3.2 is unaffected by including the intercept. However, running MLE with no intercept implicitly imposes the restriction $\Pr\left(X_{2\ell}^B = 1\right) = 1/2$, which may not hold in population. The influence function $\varphi_{\ell,j}^B(\theta_\ell)$ is the sub-vector consisting of first d_3 elements in

$$\varphi_{\ell,j}^B(\gamma_\ell) = -E\left\{H_{\ell,j}(\gamma_\ell)\right\}^{-1} S_{\ell,j}(\gamma_\ell), \tag{B2}$$

where $H_{\ell,j}(\gamma_\ell) := \partial^2 \log f_{\ell,j}(\gamma_\ell) / \left(\partial\gamma_\ell \partial\gamma_\ell^\top\right)$ is the Hessian, $S_{\ell,j}(\gamma_\ell) := \partial \log f_{\ell,j}(\gamma_\ell)/\partial\gamma_\ell$ is the score, and $\log f_{\ell,j}(\gamma_\ell)$ is the log-likelihood contribution of the j-th observation.

The Hessian and score for Probit are

$$\begin{aligned}
H_{\ell,j}^P(\gamma_\ell) = -\phi\left(\mathcal{Z}_{\ell,j}^B\right) \Bigg[& X_{2\ell,j}^B \frac{\phi\left(\mathcal{Z}_{\ell,j}^B\right) + \mathcal{Z}_{\ell,j}^B \Phi\left(\mathcal{Z}_{\ell,j}^B\right)}{\Phi^2\left(\mathcal{Z}_{\ell,j}^B\right)} \\
& + \left(1 - X_{2\ell,j}^B\right) \frac{\phi\left(\mathcal{Z}_{\ell,j}^B\right) - \mathcal{Z}_{\ell,j}^B \left\{1 - \Phi\left(\mathcal{Z}_{\ell,j}^B\right)\right\}}{\left\{1 - \Phi\left(\mathcal{Z}_{\ell,j}^B\right)\right\}^2} \Bigg] \mathcal{X}_{3j}^B \mathcal{X}_{3j}^{B\top}, \text{ and} \\
S_{\ell,j}^P(\gamma_\ell) = \phi\left(\mathcal{Z}_{\ell,j}^B\right) & \left\{ \frac{X_{2\ell,j}^B}{\Phi\left(\mathcal{Z}_{\ell,j}^B\right)} - \frac{1 - X_{2\ell,j}^B}{1 - \Phi\left(\mathcal{Z}_{\ell,j}^B\right)} \right\} \mathcal{X}_{3j}^B.
\end{aligned}$$

Those for Logit are

$$\begin{aligned}
H_{\ell,j}^L(\gamma_\ell) &= -\left[\frac{\exp\left(-\mathcal{Z}_{\ell,j}^B\right)}{\left\{1 + \exp\left(-\mathcal{Z}_{\ell,j}^B\right)\right\}^2} \right] \mathcal{X}_{3j}^B \mathcal{X}_{3j}^{B\top}, \text{ and} \\
S_{\ell,j}^L(\gamma_\ell) &= \left\{ \frac{X_{2\ell,j}^B}{1 + \exp\left(\mathcal{Z}_{\ell,j}^B\right)} - \frac{1 - X_{2\ell,j}^B}{1 + \exp\left(-\mathcal{Z}_{\ell,j}^B\right)} \right\} \mathcal{X}_{3j}^B.
\end{aligned}$$

In practice, the expectation in (B2) is replaced by its sample counterpart. The index coefficient γ_ℓ is also replaced by the corresponding MLE (denoted as $\hat{\gamma}_\ell^P$ and $\hat{\gamma}_\ell^L$ for Probit and Logit, respectively) so that the estimated index becomes $\hat{\mathcal{Z}}_{\ell,j}^B = \hat{\gamma}_\ell^{\sharp\top} \mathcal{X}_{3j}^B$ for $\sharp \in \{P, L\}$.

B.2.3 Alternative Ways of Estimating $g_{3\ell}^s(z)$

There are two other estimation methods of $g_{3\ell}^s(z)$ for $s \in \{A, B\}$ than the OLS-based method discussed above. The first method is to estimate it nonparametrically by the NW regression smoothing. Nonparametric regression estimation is robust to misspecification and particularly effective if we are uncertain about validity of the linearity condition. In contrast, the second method relies on Eq. (10) of the main text. Assuming that the marginal distribution of X_3^s is elliptically symmetric, we can estimate $g_{3\ell}^s(z)$ by

$$\hat{g}_{3\ell}^s(z) = \hat{\mu}^s + \frac{z - \hat{\theta}_\ell^\top \hat{\mu}^s}{\hat{\theta}_\ell^\top \hat{\Sigma}^s \hat{\theta}_\ell} \hat{\Sigma}^s \hat{\theta}_\ell,$$

where $\hat{\mu}^s$ and $\hat{\Sigma}^s$ are the sample mean and covariance of X_3^s. However, our simulations indicate that our performance measures ($MedSE$ and CR reported in Table D1 of Supplement D) of these alternative methods are not much different from those of OLS, and thus they are not reported.

Supplement C: Difficulties in Extending MSII

C.1 MSII: A Review

MSII proposed by Hirukawa and Prokhorov (2018) (henceforth, HP18) is yet another two-sample regression estimator. To impute the missing regressor X_2^A in Eq. (2) of the main text, HP18 construct a combined dataset via the nearest neighbor matching (NNM) with respect to two sets of overlapping variables X_3^A and X_3^B. However, estimating the regression (given in Eq. (2) of the main text) by OLS after replacing the missing X_3^A by its proxy X_3^B, called the matched-sample OLS (MSOLS), turns out to be inconsistent. This result extends the case of simple regression models analyzed by Neter et al. (1965). Inconsistency of MSOLS comes from the non-vanishing attenuation bias, which is attributed to imputing X_2^B as a proxy for the true (but missing) regressor X_2^A. The fundamental part of the bias term is the covariance matrix of the error term in the reduced form of X_2^B.

However, it is possible to analytically correct the bias generated by direct imputation of X_2^B via NNM. MSII can be recognized as a bias-corrected version of MSOLS. An appealing feature of MSII is that it is a smoothing-free estimator, and thus we do not need to select user-specified parameters including the bandwidth. In particular, HP18 estimate the error covariance matrix as a part of the non-vanishing bias in MSOLS by a difference-based variance estimation method (e.g., Horowitz and Spokoiny, 2001). This method permits us to estimate the covariance matrix directly, instead of estimating the conditional expectation $E\left(X_2^B \middle| X_3^B\right)$ nonparametrically and then calculating residuals.

Finally, MSII is available on Stata. The command `msreg` enables us to implement MSII; see Hirukawa et al. (2021) for more details.

C.2 Complicated Aspects of MSII

Once $E\left(X_2^B \middle| X_3^B\right)$ is specified as a set of single-index models (SIMs), the following technical difficulties arise. In our view, some of these difficulties cannot be immediately resolved, and we decided not to pursue the extension of MSII in parallel to PILS.

C.2.1 NNM with Respect to Estimated Indices

The NNM should be made with respect to estimated indices. The index-based matching can be viewed a special case of matching with respect to estimated propensity scores having an identity link. Abadie and Imbens (2016) and Yang and Kim (2020) develop analytical tools for propensity score matching when the index coefficients are estimated by MLE. Suitable modifications of their likelihood-based methods could work for the asymptotic analysis in this stage.

C.2.2 Equation-by-Equation NNM

The index-based matching should be made in an equation-by-equation manner, and this procedure brings additional difficulty. Suppose that as in the original MSII, we select first K matches of $X_{2\ell}^B$ for the ℓ-th missing regressor $X_{2\ell}^A$ for $\ell \in \{1, \ldots, d_2\}$ from $\mathcal{S}^B$ through the index-based matching. Let $\mathcal{J}_{\ell,K}(i) = \{j_{\ell,1}(i), \ldots, j_{\ell,K}(i)\}$ denote the index set of such K matches chosen from $\mathcal{S}^B$ for the unit i (i.e., $X_{2\ell i}^A$) in $\mathcal{S}^A$. The problem is that it is uncertain whether $\mathcal{J}_{\ell,K}(i)$ coincides with $\mathcal{J}_{\ell',K}(i)$ for $\ell \neq \ell'$. This is in sharp contrast to NNM with respect to original overlapping variables in the original MSII, which finds first K matches for the unit i (i.e., the set of all missing regressors X_{2i}^A) in $\mathcal{S}^A$ at once. Therefore, $\mathcal{J}_{\ell,K}(i) = \mathcal{J}_{\ell',K}(i)$ is always ensured.

What happens if $\mathcal{J}_{\ell,K}(i) \neq \mathcal{J}_{\ell',K}(i)$? We must monitor which (or how many) indices are chosen commonly across $\mathcal{J}_{\ell,K}(i)$ and $\mathcal{J}_{\ell',K}(i)$ to estimate off-diagonal elements of $E\left(\eta_{2(i)}^B \eta_{2(i)}^{B\top}\right)$, where

$$\eta_{2(i)}^B := \left(\frac{1}{K}\sum_{j \in \mathcal{J}_{1,K}(i)} \eta_{21,j}^B, \ldots, \frac{1}{K}\sum_{j \in \mathcal{J}_{d_2,K}(i)} \eta_{2d_2,j}^B\right)^\top.$$

It appears that this additional requirement forces us to give up the difference-based variance estimation and shift to a residual-based one. Alternatively, to preserve the smoothing-free nature of MSII, we may have to impose the stringent assumption that all off-diagonal elements of $E\left(\eta_{2(i)}^B \eta_{2(i)}^{B\top}\right)$ are zeros.

C.2.3 Ordering with Respect to Estimated Indices

Suppose that we adopt the assumption that $E\left(\eta_{2(i)}^{B}\eta_{2(i)}^{B\top}\right)$ is diagonal. The problem that arises in this stage is that difference-based variance estimation requires reordering of estimated indices. To the best of our knowledge, there is no analytical tool that can justify the index-based reordering. In fact, Kulasekera and Lin (2010) raise difficulties in analyzing difference-based variance estimation of the error term in SIMs.

Supplement D: Complete Set of Simulations

D.1 Monte Carlo Design

The simulation design is as described in Section 4.1 of the main text.

D.2 Results

The complete set of simulation results for Section 4.2 of the main text is presented in Table D1. These results use the sample sizes $(n, m) = (2000, 1000)$ and are based on 1000 Monte Carlo replications.

In addition, we provide the results for the sample sizes $(n, m) = (1000, 1000)$, based on the same number of replications, in Table D2. The results are very consistent for the two different sample size pairs.

Table D1: Monte Carlo Results for $(n, m) = (2000, 1000)$

Panel A: Distribution 1 (elliptical) - Model A: $g_{21}(z) = z$

Estimator		*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$						
For β_{21}:						
OLS*		1.0004	0.0152	0.0152	0.0151	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9415	0.1549	0.1656	0.0476	90%
PILS-SDR:						
	(SIR, SIR)	0.9320	0.1708	0.1838	0.0640	90%
	(SIR, Probit)	0.9337	0.1703	0.1827	0.0646	91%
	(SIR, Logit)	0.9340	0.1702	0.1825	0.0703	93%
	(PLS, SIR)	0.9405	0.1530	0.1641	0.0578	91%
	(PLS, Probit)	0.9422	0.1524	0.1630	0.0586	92%
	(PLS, Logit)	0.9426	0.1524	0.1628	0.0651	94%
	(PIR, SIR)	0.9316	0.1716	0.1847	0.0642	90%
	(PIR, Probit)	0.9333	0.1711	0.1836	0.0648	91%
	(PIR, Logit)	0.9336	0.1710	0.1834	0.0705	93%
For β_{31}:						
OLS*		0.9979	0.0356	0.0357	0.0359	94%
OLS-S		2.0559	0.0709	1.0583	0.0687	–
PILS-SDR*		1.0163	0.0697	0.0716	0.0548	92%
PILS-SDR:						
	(SIR, SIR)	1.0207	0.0889	0.0912	0.0820	94%
	(SIR, Probit)	1.0213	0.0889	0.0914	0.0830	95%
	(SIR, Logit)	1.0213	0.0891	0.0916	0.0901	97%
	(PLS, SIR)	1.0182	0.0840	0.0859	0.0719	94%
	(PLS, Probit)	1.0189	0.0845	0.0866	0.0730	95%
	(PLS, Logit)	1.0189	0.0847	0.0868	0.0803	96%
	(PIR, SIR)	1.0209	0.0889	0.0913	0.0823	94%
	(PIR, Probit)	1.0215	0.0889	0.0915	0.0835	95%
	(PIR, Logit)	1.0215	0.0891	0.0917	0.0905	97%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)\,,\ (\rho_N^B, \rho_D^B) = (-0.40, -0.40)$						
For β_{21}:						
OLS*		1.0004	0.0152	0.0152	0.0151	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9440	0.1505	0.1605	0.0472	90%
PILS-SDR:						
	(SIR, SIR)	0.9398	0.1500	0.1617	0.0712	92%
	(SIR, Probit)	0.9413	0.1496	0.1607	0.0719	92%
	(SIR, Logit)	0.9416	0.1496	0.1605	0.0801	94%
	(PLS, SIR)	0.9398	0.1517	0.1632	0.0620	89%
	(PLS, Probit)	0.9414	0.1514	0.1623	0.0631	90%
	(PLS, Logit)	0.9417	0.1514	0.1622	0.0716	93%
	(PIR, SIR)	0.9394	0.1509	0.1626	0.0714	92%
	(PIR, Probit)	0.9409	0.1504	0.1616	0.0720	92%
	(PIR, Logit)	0.9412	0.1504	0.1615	0.0804	95%
For β_{31}:						
OLS*		0.9979	0.0356	0.0357	0.0359	94%
OLS-S		2.0559	0.0709	1.0583	0.0687	–
PILS-SDR*		1.0147	0.0664	0.0680	0.0549	92%
PILS-SDR:						
	(SIR, SIR)	1.0180	0.0986	0.1002	0.1032	96%
	(SIR, Probit)	1.0183	0.0977	0.0994	0.1056	96%
	(SIR, Logit)	1.0185	0.0977	0.0994	0.1194	99%
	(PLS, SIR)	1.0187	0.0973	0.0991	0.0871	93%
	(PLS, Probit)	1.0190	0.0964	0.0983	0.0897	94%
	(PLS, Logit)	1.0192	0.0965	0.0984	0.1026	97%
	(PIR, SIR)	1.0182	0.0988	0.1005	0.1034	96%
	(PIR, Probit)	1.0185	0.0978	0.0996	0.1058	95%
	(PIR, Logit)	1.0187	0.0978	0.0996	0.1197	99%

Table D1: (continued)

Panel B: Distribution 1 (elliptical) - Model B: $g_{21}(z) = z + 9\phi(z)$

Estimator		*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$						
For β_{21}:						
OLS*		1.0000	0.0115	0.0115	0.0116	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9767	0.1126	0.1150	0.0301	92%
PILS-SDR:						
	(SIR, SIR)	0.9745	0.1275	0.1300	0.0361	95%
	(SIR, Probit)	0.9750	0.1273	0.1298	0.0362	95%
	(SIR, Logit)	0.9751	0.1274	0.1298	0.0370	95%
	(PLS, SIR)	0.9788	0.1136	0.1156	0.0348	95%
	(PLS, Probit)	0.9792	0.1135	0.1153	0.0348	95%
	(PLS, Logit)	0.9793	0.1135	0.1153	0.0359	95%
	(PIR, SIR)	0.9737	0.1307	0.1333	0.0362	94%
	(PIR, Probit)	0.9741	0.1306	0.1331	0.0362	95%
	(PIR, Logit)	0.9742	0.1306	0.1331	0.0370	95%
For β_{31}:						
OLS*		0.9982	0.0348	0.0349	0.0354	95%
OLS-S		2.0548	0.0820	1.0580	0.0811	–
PILS-SDR*		1.0061	0.0632	0.0635	0.0544	93%
PILS-SDR:						
	(SIR, SIR)	1.0091	0.0905	0.0910	0.0949	97%
	(SIR, Probit)	1.0098	0.0905	0.0911	0.0945	97%
	(SIR, Logit)	1.0098	0.0906	0.0912	0.0998	97%
	(PLS, SIR)	1.0106	0.0931	0.0937	0.0878	95%
	(PLS, Probit)	1.0113	0.0931	0.0937	0.0868	95%
	(PLS, Logit)	1.0114	0.0932	0.0939	0.0917	96%
	(PIR, SIR)	1.0096	0.0913	0.0918	0.0955	97%
	(PIR, Probit)	1.0103	0.0913	0.0918	0.0952	97%
	(PIR, Logit)	1.0103	0.0914	0.0920	0.1004	98%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)$, $(\rho_N^B, \rho_D^B) = (-0.40, -0.40)$						
For β_{21}:						
OLS*		1.0000	0.0115	0.0115	0.0116	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9767	0.1173	0.1196	0.0300	93%
PILS-SDR:						
	(SIR, SIR)	0.9712	0.1095	0.1132	0.0396	94%
	(SIR, Probit)	0.9717	0.1093	0.1129	0.0396	95%
	(SIR, Logit)	0.9718	0.1093	0.1129	0.0413	96%
	(PLS, SIR)	0.9630	0.1351	0.1400	0.0378	92%
	(PLS, Probit)	0.9635	0.1349	0.1398	0.0378	92%
	(PLS, Logit)	0.9636	0.1349	0.1398	0.0394	93%
	(PIR, SIR)	0.9713	0.1095	0.1132	0.0397	95%
	(PIR, Probit)	0.9717	0.1093	0.1129	0.0397	95%
	(PIR, Logit)	0.9718	0.1093	0.1129	0.0413	96%
For β_{31}:						
OLS*		0.9982	0.0348	0.0349	0.0354	95%
OLS-S		2.0548	0.0820	1.0580	0.0811	–
PILS-SDR*		1.0051	0.0614	0.0616	0.0544	94%
PILS-SDR:						
	(SIR, SIR)	1.0087	0.1121	0.1124	0.1252	97%
	(SIR, Probit)	1.0093	0.1114	0.1118	0.1250	97%
	(SIR, Logit)	1.0095	0.1113	0.1117	0.1350	98%
	(PLS, SIR)	1.0134	0.1199	0.1207	0.1129	94%
	(PLS, Probit)	1.0143	0.1199	0.1207	0.1120	94%
	(PLS, Logit)	1.0145	0.1199	0.1207	0.1212	96%
	(PIR, SIR)	1.0087	0.1121	0.1124	0.1254	97%
	(PIR, Probit)	1.0093	0.1114	0.1118	0.1251	97%
	(PIR, Logit)	1.0095	0.1113	0.1117	0.1351	98%

Table D1: (continued)

Panel C: Distribution 1 (elliptical) - Model C: $g_{21}(z) = [2\mathbb{I}\{z > 0\} + 1]\sqrt{|z|}$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$					
For β_{21}:					
OLS*	1.0004	0.0148	0.0148	0.0145	94%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9760	0.1278	0.1301	0.0437	94%
PILS-SDR:					
(SIR, SIR)	0.9699	0.1328	0.1362	0.0515	94%
(SIR, Probit)	0.9706	0.1326	0.1359	0.0517	94%
(SIR, Logit)	0.9707	0.1326	0.1358	0.0537	95%
(PLS, SIR)	0.9718	0.1299	0.1329	0.0509	93%
(PLS, Probit)	0.9725	0.1297	0.1326	0.0512	93%
(PLS, Logit)	0.9727	0.1297	0.1326	0.0535	94%
(PIR, SIR)	0.9694	0.1349	0.1383	0.0515	94%
(PIR, Probit)	0.9701	0.1347	0.1380	0.0517	94%
(PIR, Logit)	0.9702	0.1347	0.1380	0.0538	94%
For β_{31}:					
OLS*	0.9980	0.0353	0.0354	0.0352	95%
OLS-S	1.7972	0.0709	0.8004	0.0680	–
PILS-SDR*	1.0033	0.0595	0.0596	0.0545	94%
PILS-SDR:					
(SIR, SIR)	1.0066	0.0777	0.0779	0.0815	97%
(SIR, Probit)	1.0072	0.0777	0.0780	0.0835	97%
(SIR, Logit)	1.0072	0.0777	0.0780	0.0905	99%
(PLS, SIR)	1.0063	0.0827	0.0829	0.0792	94%
(PLS, Probit)	1.0069	0.0825	0.0828	0.0813	95%
(PLS, Logit)	1.0069	0.0825	0.0828	0.0874	97%
(PIR, SIR)	1.0067	0.0777	0.0779	0.0818	97%
(PIR, Probit)	1.0073	0.0777	0.0780	0.0839	97%
(PIR, Logit)	1.0073	0.0777	0.0780	0.0907	99%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)$, $(\rho_N^B, \rho_D^B) = (-0.40, -0.40)$					
For β_{21}:					
OLS*	1.0004	0.0148	0.0148	0.0145	94%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9760	0.1369	0.1390	0.0436	93%
PILS-SDR:					
(SIR, SIR)	0.9598	0.1317	0.1377	0.0545	92%
(SIR, Probit)	0.9604	0.1315	0.1373	0.0551	93%
(SIR, Logit)	0.9606	0.1315	0.1373	0.0582	94%
(PLS, SIR)	0.9514	0.1445	0.1525	0.0535	89%
(PLS, Probit)	0.9520	0.1443	0.1521	0.0542	90%
(PLS, Logit)	0.9522	0.1444	0.1521	0.0577	92%
(PIR, SIR)	0.9601	0.1315	0.1374	0.0546	92%
(PIR, Probit)	0.9607	0.1313	0.1370	0.0552	93%
(PIR, Logit)	0.9609	0.1313	0.1370	0.0582	94%
For β_{31}:					
OLS*	0.9980	0.0353	0.0354	0.0352	95%
OLS-S	1.7972	0.0709	0.8004	0.0680	–
PILS-SDR*	1.0031	0.0591	0.0591	0.0545	93%
PILS-SDR:					
(SIR, SIR)	1.0094	0.0931	0.0936	0.1044	97%
(SIR, Probit)	1.0099	0.0925	0.0930	0.1077	98%
(SIR, Logit)	1.0101	0.0926	0.0931	0.1210	99%
(PLS, SIR)	1.0106	0.1000	0.1005	0.0989	95%
(PLS, Probit)	1.0111	0.0990	0.0996	0.1025	95%
(PLS, Logit)	1.0113	0.0990	0.0997	0.1142	97%
(PIR, SIR)	1.0094	0.0933	0.0938	0.1045	97%
(PIR, Probit)	1.0099	0.0927	0.0932	0.1077	98%
(PIR, Logit)	1.0101	0.0928	0.0933	0.1209	99%

Table D1: (continued)

Panel D: Distribution 2 (non-elliptical) - Model A: $g_{21}(z) = z$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_U^A, \rho_D^A) = (\rho_U^B, \rho_D^B) = (0.35, 0.40)$					
For β_{21}:					
OLS*	0.9995	0.0136	0.0136	0.0137	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9415	0.1543	0.1650	0.0427	90%
PILS-SDR:					
(SIR, SIR)	0.9482	0.1403	0.1496	0.0564	93%
(SIR, Probit)	0.9477	0.1401	0.1496	0.0563	93%
(SIR, Logit)	0.9479	0.1402	0.1496	0.0594	94%
(PLS, SIR)	0.9496	0.1393	0.1481	0.0579	93%
(PLS, Probit)	0.9492	0.1387	0.1477	0.0582	93%
(PLS, Logit)	0.9495	0.1391	0.1480	0.0610	94%
(PIR, SIR)	0.9467	0.1463	0.1557	0.0606	93%
(PIR, Probit)	0.9461	0.1461	0.1557	0.0610	93%
(PIR, Logit)	0.9464	0.1462	0.1557	0.0637	95%
For β_{31}:					
OLS*	1.0004	0.0329	0.0329	0.0328	96%
OLS-S	2.1186	0.0710	1.1208	0.0707	–
PILS-SDR*	1.0180	0.0633	0.0658	0.0494	91%
PILS-SDR:					
(SIR, SIR)	1.0152	0.0736	0.0751	0.0735	96%
(SIR, Probit)	1.0165	0.0731	0.0749	0.0738	96%
(SIR, Logit)	1.0166	0.0732	0.0750	0.0793	98%
(PLS, SIR)	1.0158	0.0707	0.0724	0.0761	96%
(PLS, Probit)	1.0173	0.0707	0.0727	0.0765	96%
(PLS, Logit)	1.0173	0.0706	0.0727	0.0825	98%
(PIR, SIR)	1.0155	0.0740	0.0756	0.0816	97%
(PIR, Probit)	1.0169	0.0736	0.0756	0.0818	97%
(PIR, Logit)	1.0170	0.0737	0.0756	0.0874	98%
Heterogeneous Populations: $(\rho_U^A, \rho_D^A) = (0.35, 0.40)$, $(\rho_U^B, \rho_D^B) = (-0.35, -0.40)$					
For β_{21}:					
OLS*	0.9995	0.0136	0.0136	0.0137	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9535	0.1280	0.1362	0.0428	90%
PILS-SDR:					
(SIR, SIR)	0.9428	0.1582	0.1683	0.0602	93%
(SIR, Probit)	0.9422	0.1579	0.1682	0.0605	92%
(SIR, Logit)	0.9426	0.1578	0.1679	0.0639	94%
(PLS, SIR)	0.9468	0.1475	0.1568	0.0571	92%
(PLS, Probit)	0.9465	0.1474	0.1568	0.0577	93%
(PLS, Logit)	0.9469	0.1473	0.1566	0.0617	94%
(PIR, SIR)	0.9425	0.1585	0.1686	0.0634	93%
(PIR, Probit)	0.9420	0.1581	0.1684	0.0635	93%
(PIR, Logit)	0.9423	0.1580	0.1682	0.0671	94%
For β_{31}:					
OLS*	1.0004	0.0329	0.0329	0.0328	96%
OLS-S	2.1186	0.0710	1.1208	0.0707	–
PILS-SDR*	1.0149	0.0593	0.0611	0.0495	92%
PILS-SDR:					
(SIR, SIR)	1.0169	0.0813	0.0830	0.0788	95%
(SIR, Probit)	1.0179	0.0807	0.0827	0.0792	95%
(SIR, Logit)	1.0181	0.0809	0.0829	0.0854	97%
(PLS, SIR)	1.0155	0.0746	0.0762	0.0747	96%
(PLS, Probit)	1.0167	0.0749	0.0768	0.0755	96%
(PLS, Logit)	1.0169	0.0749	0.0768	0.0832	98%
(PIR, SIR)	1.0168	0.0811	0.0828	0.0849	96%
(PIR, Probit)	1.0178	0.0804	0.0823	0.0858	96%
(PIR, Logit)	1.0180	0.0805	0.0825	0.0921	98%

Table D1: (continued)

Panel E: Distribution 2 (non-elliptical) - Model B: $g_{21}(z) = z + 9\phi(z)$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_U^A, \rho_D^A) = (\rho_U^B, \rho_D^B) = (0.35, 0.40)$					
For β_{21}:					
OLS*	0.9997	0.0107	0.0107	0.0107	96%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9769	0.1171	0.1194	0.0281	94%
PILS-SDR:					
(SIR, SIR)	0.9779	0.1196	0.1216	0.0349	95%
(SIR, Probit)	0.9777	0.1195	0.1215	0.0348	95%
(SIR, Logit)	0.9777	0.1195	0.1215	0.0353	95%
(PLS, SIR)	0.9817	0.1103	0.1118	0.0346	95%
(PLS, Probit)	0.9815	0.1102	0.1117	0.0345	95%
(PLS, Logit)	0.9816	0.1102	0.1117	0.0350	95%
(PIR, SIR)	0.9818	0.1048	0.1064	0.0361	96%
(PIR, Probit)	0.9815	0.1047	0.1063	0.0359	96%
(PIR, Logit)	0.9816	0.1047	0.1063	0.0364	96%
For β_{31}:					
OLS*	1.0003	0.0325	0.0325	0.0323	96%
OLS-S	2.1193	0.0833	1.1224	0.0813	–
PILS-SDR*	1.0079	0.0571	0.0577	0.0493	94%
PILS-SDR:					
(SIR, SIR)	1.0115	0.0807	0.0815	0.0878	98%
(SIR, Probit)	1.0127	0.0804	0.0814	0.0863	97%
(SIR, Logit)	1.0129	0.0805	0.0816	0.0893	98%
(PLS, SIR)	1.0116	0.0816	0.0824	0.0860	97%
(PLS, Probit)	1.0128	0.0812	0.0822	0.0844	96%
(PLS, Logit)	1.0129	0.0812	0.0822	0.0878	97%
(PIR, SIR)	1.0105	0.0783	0.0790	0.0970	98%
(PIR, Probit)	1.0116	0.0779	0.0788	0.0957	98%
(PIR, Logit)	1.0117	0.0780	0.0789	0.0997	98%
Heterogeneous Populations: $(\rho_U^A, \rho_D^A) = (0.35, 0.40)$, $(\rho_U^B, \rho_D^B) = (-0.35, -0.40)$					
For β_{21}:					
OLS*	0.9997	0.0107	0.0107	0.0107	96%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9854	0.0970	0.0981	0.0280	94%
PILS-SDR:					
(SIR, SIR)	0.9780	0.1151	0.1172	0.0363	95%
(SIR, Probit)	0.9778	0.1150	0.1171	0.0361	94%
(SIR, Logit)	0.9779	0.1149	0.1170	0.0369	95%
(PLS, SIR)	0.9837	0.0976	0.0990	0.0358	95%
(PLS, Probit)	0.9835	0.0976	0.0989	0.0356	95%
(PLS, Logit)	0.9836	0.0975	0.0989	0.0364	95%
(PIR, SIR)	0.9778	0.1158	0.1179	0.0374	95%
(PIR, Probit)	0.9776	0.1156	0.1178	0.0372	95%
(PIR, Logit)	0.9777	0.1156	0.1177	0.0380	95%
For β_{31}:					
OLS*	1.0003	0.0325	0.0325	0.0323	96%
OLS-S	2.1193	0.0833	1.1224	0.0813	–
PILS-SDR*	1.0060	0.0560	0.0563	0.0494	94%
PILS-SDR:					
(SIR, SIR)	1.0068	0.0838	0.0841	0.0945	97%
(SIR, Probit)	1.0077	0.0831	0.0834	0.0932	98%
(SIR, Logit)	1.0079	0.0831	0.0835	0.0973	98%
(PLS, SIR)	1.0070	0.0824	0.0827	0.0909	97%
(PLS, Probit)	1.0079	0.0817	0.0821	0.0898	97%
(PLS, Logit)	1.0081	0.0817	0.0821	0.0945	98%
(PIR, SIR)	1.0065	0.0833	0.0836	0.1031	98%
(PIR, Probit)	1.0075	0.0826	0.0829	0.1019	98%
(PIR, Logit)	1.0076	0.0826	0.0830	0.1062	99%

Table D1: (continued)

Panel F: Distribution 2 (non-elliptical) - Model C: $g_{21}(z) = [2\mathbb{I}\{z > 0\} + 1]\sqrt{|z|}$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_U^A, \rho_D^A) = (\rho_U^B, \rho_D^B) = (0.35, 0.40)$					
For β_{21}:					
OLS*	0.9994	0.0137	0.0138	0.0138	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9742	0.1356	0.1381	0.0418	92%
PILS-SDR:					
(SIR, SIR)	0.9742	0.1280	0.1306	0.0487	94%
(SIR, Probit)	0.9738	0.1277	0.1304	0.0490	93%
(SIR, Logit)	0.9739	0.1279	0.1305	0.0503	94%
(PLS, SIR)	0.9710	0.1353	0.1384	0.0490	93%
(PLS, Probit)	0.9704	0.1351	0.1383	0.0493	93%
(PLS, Logit)	0.9704	0.1351	0.1383	0.0508	94%
(PIR, SIR)	0.9742	0.1272	0.1298	0.0504	94%
(PIR, Probit)	0.9738	0.1269	0.1296	0.0508	94%
(PIR, Logit)	0.9738	0.1270	0.1297	0.0518	95%
For β_{31}:					
OLS*	1.0004	0.0318	0.0318	0.0321	96%
OLS-S	1.7696	0.0646	0.7723	0.0659	–
PILS-SDR*	1.0056	0.0526	0.0529	0.0494	94%
PILS-SDR:					
(SIR, SIR)	1.0032	0.0666	0.0666	0.0744	97%
(SIR, Probit)	1.0044	0.0659	0.0660	0.0758	98%
(SIR, Logit)	1.0045	0.0659	0.0661	0.0810	99%
(PLS, SIR)	1.0014	0.0709	0.0709	0.0741	97%
(PLS, Probit)	1.0024	0.0701	0.0701	0.0757	97%
(PLS, Logit)	1.0026	0.0702	0.0702	0.0814	98%
(PIR, SIR)	1.0034	0.0667	0.0668	0.0790	98%
(PIR, Probit)	1.0045	0.0658	0.0659	0.0806	98%
(PIR, Logit)	1.0047	0.0659	0.0661	0.0861	99%
Heterogeneous Populations: $(\rho_U^A, \rho_D^A) = (0.35, 0.40)$, $(\rho_U^B, \rho_D^B) = (-0.35, -0.40)$					
For β_{21}:					
OLS*	0.9994	0.0137	0.0138	0.0138	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9875	0.1047	0.1055	0.0417	94%
PILS-SDR:					
(SIR, SIR)	0.9697	0.1252	0.1288	0.0500	95%
(SIR, Probit)	0.9693	0.1249	0.1287	0.0504	95%
(SIR, Logit)	0.9694	0.1249	0.1286	0.0521	96%
(PLS, SIR)	0.9706	0.1230	0.1265	0.0501	93%
(PLS, Probit)	0.9701	0.1227	0.1263	0.0504	94%
(PLS, Logit)	0.9702	0.1227	0.1263	0.0520	95%
(PIR, SIR)	0.9698	0.1223	0.1260	0.0516	95%
(PIR, Probit)	0.9695	0.1220	0.1258	0.0518	95%
(PIR, Logit)	0.9696	0.1220	0.1258	0.0535	95%
For β_{31}:					
OLS*	1.0004	0.0318	0.0318	0.0321	96%
OLS-S	1.7696	0.0646	0.7723	0.0659	–
PILS-SDR*	1.0038	0.0502	0.0503	0.0494	95%
PILS-SDR:					
(SIR, SIR)	1.0056	0.0696	0.0699	0.0789	98%
(SIR, Probit)	1.0065	0.0690	0.0693	0.0807	98%
(SIR, Logit)	1.0067	0.0691	0.0695	0.0872	99%
(PLS, SIR)	1.0036	0.0734	0.0735	0.0776	97%
(PLS, Probit)	1.0044	0.0727	0.0728	0.0794	97%
(PLS, Logit)	1.0046	0.0727	0.0728	0.0861	98%
(PIR, SIR)	1.0058	0.0694	0.0697	0.0835	98%
(PIR, Probit)	1.0067	0.0690	0.0693	0.0848	98%
(PIR, Logit)	1.0069	0.0690	0.0694	0.0918	99%

Notes: *Mean* = simulation average of the parameter estimate; *SD* = simulation standard deviation of the parameter estimate; *RMSE* = root mean-squared error of the parameter estimate; *MedSE* = simulation median of the standard error; and *CR* = coverage rate against the nominal 95% confidence interval.

Table D2: Monte Carlo Results for $(n, m) = (1000, 1000)$

Panel A: Distribution 1 (elliptical) - Model A: $g_{21}(z) = z$

Estimator		*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$						
For β_{21}:						
OLS*		0.9999	0.0216	0.0216	0.0214	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9510	0.1488	0.1567	0.0620	91%
PILS-SDR:						
	(SIR, SIR)	0.9512	0.1340	0.1426	0.0756	92%
	(SIR, Probit)	0.9525	0.1336	0.1418	0.0759	92%
	(SIR, Logit)	0.9529	0.1334	0.1414	0.0806	95%
	(PLS, SIR)	0.9535	0.1327	0.1406	0.0705	90%
	(PLS, Probit)	0.9549	0.1323	0.1398	0.0710	91%
	(PLS, Logit)	0.9553	0.1322	0.1396	0.0763	93%
	(PIR, SIR)	0.9513	0.1337	0.1423	0.0757	92%
	(PIR, Probit)	0.9526	0.1333	0.1415	0.0762	92%
	(PIR, Logit)	0.9530	0.1331	0.1411	0.0808	95%
For β_{31}:						
OLS*		1.0001	0.0517	0.0517	0.0509	95%
OLS-S		2.0582	0.0961	1.0625	0.0970	–
PILS-SDR*		1.0163	0.0876	0.0891	0.0776	93%
PILS-SDR:						
	(SIR, SIR)	1.0191	0.0958	0.0977	0.0985	96%
	(SIR, Probit)	1.0191	0.0956	0.0974	0.0994	96%
	(SIR, Logit)	1.0192	0.0957	0.0976	0.1055	97%
	(PLS, SIR)	1.0183	0.0953	0.0970	0.0906	94%
	(PLS, Probit)	1.0183	0.0950	0.0968	0.0916	95%
	(PLS, Logit)	1.0184	0.0953	0.0970	0.0979	96%
	(PIR, SIR)	1.0191	0.0957	0.0976	0.0988	96%
	(PIR, Probit)	1.0191	0.0954	0.0973	0.0998	96%
	(PIR, Logit)	1.0192	0.0955	0.0974	0.1059	97%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)$, $(\rho_N^B, \rho_D^B) = (-0.40, -0.40)$						
For β_{21}:						
OLS*		0.9999	0.0216	0.0216	0.0214	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9506	0.1530	0.1608	0.0621	90%
PILS-SDR:						
	(SIR, SIR)	0.9449	0.1483	0.1582	0.0810	91%
	(SIR, Probit)	0.9459	0.1480	0.1575	0.0815	92%
	(SIR, Logit)	0.9462	0.1480	0.1574	0.0888	93%
	(PLS, SIR)	0.9454	0.1437	0.1537	0.0736	90%
	(PLS, Probit)	0.9464	0.1432	0.1529	0.0745	91%
	(PLS, Logit)	0.9467	0.1432	0.1528	0.0820	93%
	(PIR, SIR)	0.9444	0.1503	0.1603	0.0811	91%
	(PIR, Probit)	0.9454	0.1500	0.1596	0.0815	92%
	(PIR, Logit)	0.9457	0.1500	0.1595	0.0888	93%
For β_{31}:						
OLS*		1.0001	0.0517	0.0517	0.0509	95%
OLS-S		2.0582	0.0961	1.0625	0.0970	–
PILS-SDR*		1.0159	0.0862	0.0876	0.0777	94%
PILS-SDR:						
	(SIR, SIR)	1.0202	0.1126	0.1144	0.1177	96%
	(SIR, Probit)	1.0207	0.1123	0.1142	0.1198	97%
	(SIR, Logit)	1.0207	0.1124	0.1143	0.1321	98%
	(PLS, SIR)	1.0197	0.1093	0.1111	0.1029	94%
	(PLS, Probit)	1.0201	0.1089	0.1107	0.1052	94%
	(PLS, Logit)	1.0202	0.1089	0.1107	0.1167	97%
	(PIR, SIR)	1.0205	0.1130	0.1149	0.1178	96%
	(PIR, Probit)	1.0209	0.1127	0.1146	0.1199	97%
	(PIR, Logit)	1.0210	0.1127	0.1147	0.1321	98%

Table D2: (continued)

Panel B: Distribution 1 (elliptical) - Model B: $g_{21}(z) = z + 9\phi(z)$

Estimator		*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$						
For β_{21}:						
OLS*		0.9995	0.0166	0.0166	0.0163	94%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9808	0.1130	0.1146	0.0372	93%
PILS-SDR:						
	(SIR, SIR)	0.9814	0.1165	0.1179	0.0430	93%
	(SIR, Probit)	0.9818	0.1164	0.1179	0.0430	93%
	(SIR, Logit)	0.9819	0.1164	0.1178	0.0437	93%
	(PLS, SIR)	0.9869	0.0916	0.0925	0.0419	93%
	(PLS, Probit)	0.9873	0.0915	0.0924	0.0419	93%
	(PLS, Logit)	0.9874	0.0915	0.0924	0.0428	93%
	(PIR, SIR)	0.9812	0.1176	0.1191	0.0430	93%
	(PIR, Probit)	0.9815	0.1176	0.1191	0.0430	93%
	(PIR, Logit)	0.9817	0.1176	0.1190	0.0437	94%
For β_{31}:						
OLS*		1.0004	0.0510	0.0510	0.0502	95%
OLS-S		2.0571	0.1133	1.0632	0.1145	–
PILS-SDR*		1.0076	0.0827	0.0830	0.0766	95%
PILS-SDR:						
	(SIR, SIR)	1.0100	0.1027	0.1032	0.1102	97%
	(SIR, Probit)	1.0103	0.1026	0.1032	0.1095	97%
	(SIR, Logit)	1.0104	0.1027	0.1033	0.1138	98%
	(PLS, SIR)	1.0107	0.1042	0.1047	0.1040	95%
	(PLS, Probit)	1.0110	0.1041	0.1047	0.1032	95%
	(PLS, Logit)	1.0112	0.1042	0.1048	0.1079	96%
	(PIR, SIR)	1.0100	0.1028	0.1033	0.1106	97%
	(PIR, Probit)	1.0103	0.1027	0.1032	0.1100	97%
	(PIR, Logit)	1.0105	0.1028	0.1033	0.1143	98%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)$, $(\rho_N^B, \rho_D^B) = (-0.40, -0.40)$						
For β_{21}:						
OLS*		0.9995	0.0166	0.0166	0.0163	94%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9787	0.1222	0.1241	0.0373	92%
PILS-SDR:						
	(SIR, SIR)	0.9748	0.1099	0.1127	0.0458	93%
	(SIR, Probit)	0.9751	0.1098	0.1126	0.0458	93%
	(SIR, Logit)	0.9752	0.1098	0.1125	0.0469	94%
	(PLS, SIR)	0.9744	0.1110	0.1140	0.0441	92%
	(PLS, Probit)	0.9746	0.1110	0.1138	0.0441	92%
	(PLS, Logit)	0.9747	0.1110	0.1138	0.0454	93%
	(PIR, SIR)	0.9744	0.1131	0.1159	0.0458	93%
	(PIR, Probit)	0.9746	0.1130	0.1158	0.0459	93%
	(PIR, Logit)	0.9747	0.1129	0.1157	0.0469	94%
For β_{31}:						
OLS*		1.0004	0.0510	0.0510	0.0502	95%
OLS-S		2.0571	0.1133	1.0632	0.1145	–
PILS-SDR*		1.0077	0.0820	0.0824	0.0768	95%
PILS-SDR:						
	(SIR, SIR)	1.0119	0.1204	0.1210	0.1372	97%
	(SIR, Probit)	1.0126	0.1204	0.1211	0.1371	97%
	(SIR, Logit)	1.0128	0.1205	0.1212	0.1468	98%
	(PLS, SIR)	1.0104	0.1254	0.1258	0.1253	95%
	(PLS, Probit)	1.0110	0.1252	0.1257	0.1248	95%
	(PLS, Logit)	1.0112	0.1252	0.1257	0.1334	96%
	(PIR, SIR)	1.0121	0.1207	0.1213	0.1373	97%
	(PIR, Probit)	1.0128	0.1208	0.1214	0.1372	97%
	(PIR, Logit)	1.0130	0.1209	0.1215	0.1469	98%

Table D2: (continued)

Panel C: Distribution 1 (elliptical) - Model C: $g_{21}(z) = [2\mathbb{I}\{z > 0\} + 1]\sqrt{|z|}$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_N^A, \rho_D^A) = (\rho_N^B, \rho_D^B) = (0.40, 0.40)$					
For β_{21}:					
OLS*	1.0001	0.0205	0.0205	0.0205	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9833	0.1290	0.1300	0.0543	93%
PILS-SDR:					
(SIR, SIR)	0.9810	0.1169	0.1184	0.0609	94%
(SIR, Probit)	0.9815	0.1168	0.1183	0.0612	94%
(SIR, Logit)	0.9817	0.1168	0.1183	0.0632	95%
(PLS, SIR)	0.9760	0.1266	0.1289	0.0605	93%
(PLS, Probit)	0.9765	0.1265	0.1287	0.0609	93%
(PLS, Logit)	0.9767	0.1265	0.1286	0.0629	94%
(PIR, SIR)	0.9809	0.1175	0.1190	0.0609	94%
(PIR, Probit)	0.9814	0.1174	0.1189	0.0612	94%
(PIR, Logit)	0.9816	0.1174	0.1189	0.0632	95%
For β_{31}:					
OLS*	1.0000	0.0508	0.0508	0.0500	94%
OLS-S	1.7999	0.0966	0.8058	0.0960	–
PILS-SDR*	1.0053	0.0792	0.0794	0.0768	95%
PILS-SDR:					
(SIR, SIR)	1.0097	0.0914	0.0919	0.0995	96%
(SIR, Probit)	1.0099	0.0910	0.0915	0.1011	96%
(SIR, Logit)	1.0100	0.0910	0.0916	0.1067	98%
(PLS, SIR)	1.0099	0.0966	0.0971	0.0968	95%
(PLS, Probit)	1.0100	0.0962	0.0967	0.0986	96%
(PLS, Logit)	1.0101	0.0963	0.0968	0.1039	97%
(PIR, SIR)	1.0097	0.0914	0.0919	0.0998	96%
(PIR, Probit)	1.0099	0.0910	0.0915	0.1013	96%
(PIR, Logit)	1.0100	0.0910	0.0916	0.1070	98%
Heterogeneous Populations: $(\rho_N^A, \rho_D^A) = (0.40, 0.40)$, $(\rho_N^B, \rho_D^B) = (-0.40, -0.40)$					
For β_{21}:					
OLS*	1.0001	0.0205	0.0205	0.0205	95%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9861	0.1288	0.1296	0.0546	94%
PILS-SDR:					
(SIR, SIR)	0.9706	0.1190	0.1226	0.0642	94%
(SIR, Probit)	0.9711	0.1189	0.1224	0.0647	94%
(SIR, Logit)	0.9712	0.1190	0.1224	0.0678	95%
(PLS, SIR)	0.9649	0.1273	0.1320	0.0631	93%
(PLS, Probit)	0.9654	0.1271	0.1317	0.0636	93%
(PLS, Logit)	0.9655	0.1272	0.1318	0.0666	94%
(PIR, SIR)	0.9707	0.1182	0.1218	0.0643	94%
(PIR, Probit)	0.9712	0.1181	0.1215	0.0647	94%
(PIR, Logit)	0.9713	0.1181	0.1216	0.0679	95%
For β_{31}:					
OLS*	1.0000	0.0508	0.0508	0.0500	94%
OLS-S	1.7999	0.0966	0.8058	0.0960	–
PILS-SDR*	1.0041	0.0776	0.0777	0.0770	96%
PILS-SDR:					
(SIR, SIR)	1.0116	0.1078	0.1084	0.1184	98%
(SIR, Probit)	1.0121	0.1072	0.1079	0.1215	98%
(SIR, Logit)	1.0123	0.1072	0.1079	0.1326	99%
(PLS, SIR)	1.0113	0.1153	0.1159	0.1140	95%
(PLS, Probit)	1.0118	0.1146	0.1152	0.1168	95%
(PLS, Logit)	1.0121	0.1147	0.1153	0.1274	98%
(PIR, SIR)	1.0116	0.1078	0.1084	0.1186	98%
(PIR, Probit)	1.0121	0.1072	0.1079	0.1215	98%
(PIR, Logit)	1.0123	0.1072	0.1079	0.1326	99%

Table D2: (continued)

Panel D: Distribution 2 (non-elliptical) - Model A: $g_{21}(z) = z$

Estimator		$Mean$	SD	$RMSE$	$MedSE$	CR
Homogeneous Populations: $(\rho_U^A, \rho_D^A) = (\rho_U^B, \rho_D^B) = (0.35, 0.40)$						
For β_{21}:						
OLS*		0.9999	0.0198	0.0198	0.0194	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9499	0.1482	0.1564	0.0551	90%
PILS-SDR:						
	(SIR, SIR)	0.9513	0.1490	0.1568	0.0664	92%
	(SIR, Probit)	0.9508	0.1486	0.1565	0.0667	93%
	(SIR, Logit)	0.9512	0.1486	0.1564	0.0693	94%
	(PLS, SIR)	0.9488	0.1500	0.1585	0.0679	93%
	(PLS, Probit)	0.9486	0.1494	0.1580	0.0680	94%
	(PLS, Logit)	0.9489	0.1494	0.1579	0.0706	94%
	(PIR, SIR)	0.9507	0.1515	0.1593	0.0702	93%
	(PIR, Probit)	0.9502	0.1510	0.1590	0.0704	93%
	(PIR, Logit)	0.9506	0.1511	0.1589	0.0726	94%
For β_{31}:						
OLS*		1.0011	0.0459	0.0459	0.0464	96%
OLS-S		2.1181	0.1024	1.1228	0.1002	–
PILS-SDR*		1.0139	0.0818	0.0830	0.0701	93%
PILS-SDR:						
	(SIR, SIR)	1.0131	0.0920	0.0930	0.0893	95%
	(SIR, Probit)	1.0147	0.0920	0.0932	0.0894	95%
	(SIR, Logit)	1.0148	0.0923	0.0934	0.0938	96%
	(PLS, SIR)	1.0141	0.0899	0.0910	0.0911	96%
	(PLS, Probit)	1.0157	0.0897	0.0911	0.0915	96%
	(PLS, Logit)	1.0158	0.0899	0.0913	0.0959	98%
	(PIR, SIR)	1.0134	0.0926	0.0935	0.0957	96%
	(PIR, Probit)	1.0150	0.0925	0.0937	0.0959	96%
	(PIR, Logit)	1.0151	0.0927	0.0939	0.1004	98%
Heterogeneous Populations: $(\rho_U^A, \rho_D^A) = (0.35, 0.40)$, $(\rho_U^B, \rho_D^B) = (-0.35, -0.40)$						
For β_{21}:						
OLS*		0.9999	0.0198	0.0198	0.0194	95%
OLS-S		–	–	–	–	–
PILS-SDR*		0.9611	0.1120	0.1186	0.0549	91%
PILS-SDR:						
	(SIR, SIR)	0.9594	0.1183	0.1250	0.0692	93%
	(SIR, Probit)	0.9592	0.1178	0.1247	0.0691	93%
	(SIR, Logit)	0.9595	0.1177	0.1245	0.0726	94%
	(PLS, SIR)	0.9564	0.1298	0.1369	0.0682	93%
	(PLS, Probit)	0.9563	0.1295	0.1366	0.0682	93%
	(PLS, Logit)	0.9567	0.1294	0.1365	0.0716	94%
	(PIR, SIR)	0.9593	0.1177	0.1245	0.0723	93%
	(PIR, Probit)	0.9590	0.1173	0.1242	0.0727	94%
	(PIR, Logit)	0.9594	0.1171	0.1240	0.0757	95%
For β_{31}:						
OLS*		1.0011	0.0459	0.0459	0.0464	96%
OLS-S		2.1181	0.1024	1.1228	0.1002	–
PILS-SDR*		1.0106	0.0747	0.0755	0.0700	94%
PILS-SDR:						
	(SIR, SIR)	1.0126	0.0898	0.0907	0.0933	96%
	(SIR, Probit)	1.0136	0.0893	0.0904	0.0938	97%
	(SIR, Logit)	1.0137	0.0894	0.0904	0.0989	97%
	(PLS, SIR)	1.0131	0.0894	0.0904	0.0926	96%
	(PLS, Probit)	1.0143	0.0895	0.0907	0.0929	96%
	(PLS, Logit)	1.0144	0.0896	0.0907	0.0989	97%
	(PIR, SIR)	1.0126	0.0896	0.0905	0.0998	97%
	(PIR, Probit)	1.0137	0.0892	0.0902	0.1003	97%
	(PIR, Logit)	1.0137	0.0892	0.0903	0.1052	98%

Table D2: (continued)

Panel E: Distribution 2 (non-elliptical) - Model B: $g_{21}(z) = z + 9\phi(z)$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $(\rho_U^A, \rho_D^A) = (\rho_U^B, \rho_D^B) = (0.35, 0.40)$					
For β_{21}:					
OLS*	1.0003	0.0158	0.0158	0.0151	94%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9814	0.1179	0.1194	0.0349	92%
PILS-SDR:					
(SIR, SIR)	0.9853	0.1156	0.1165	0.0410	94%
(SIR, Probit)	0.9850	0.1154	0.1164	0.0409	94%
(SIR, Logit)	0.9851	0.1155	0.1164	0.0412	95%
(PLS, SIR)	0.9817	0.1169	0.1183	0.0414	95%
(PLS, Probit)	0.9815	0.1168	0.1183	0.0413	95%
(PLS, Logit)	0.9816	0.1169	0.1183	0.0417	95%
(PIR, SIR)	0.9839	0.1203	0.1214	0.0424	95%
(PIR, Probit)	0.9837	0.1202	0.1213	0.0422	95%
(PIR, Logit)	0.9838	0.1202	0.1213	0.0426	95%
For β_{31}:					
OLS*	1.0008	0.0455	0.0455	0.0458	96%
OLS-S	2.1159	0.1184	1.1222	0.1152	–
PILS-SDR*	1.0049	0.0785	0.0786	0.0697	93%
PILS-SDR:					
(SIR, SIR)	1.0044	0.0963	0.0964	0.1017	97%
(SIR, Probit)	1.0057	0.0960	0.0962	0.1004	97%
(SIR, Logit)	1.0058	0.0962	0.0963	0.1031	98%
(PLS, SIR)	1.0054	0.0989	0.0991	0.1049	96%
(PLS, Probit)	1.0068	0.0985	0.0987	0.1035	96%
(PLS, Logit)	1.0069	0.0984	0.0987	0.1059	97%
(PIR, SIR)	1.0045	0.0972	0.0973	0.1108	97%
(PIR, Probit)	1.0058	0.0969	0.0971	0.1098	97%
(PIR, Logit)	1.0060	0.0971	0.0973	0.1121	98%
Heterogeneous Populations: $(\rho_U^A, \rho_D^A) = (0.35, 0.40)\,,\ (\rho_U^B, \rho_D^B) = (-0.35, -0.40)$					
For β_{21}:					
OLS*	1.0003	0.0158	0.0158	0.0151	94%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9916	0.0770	0.0774	0.0349	94%
PILS-SDR:					
(SIR, SIR)	0.9855	0.1056	0.1066	0.0421	96%
(SIR, Probit)	0.9853	0.1057	0.1067	0.0419	95%
(SIR, Logit)	0.9854	0.1057	0.1067	0.0425	95%
(PLS, SIR)	0.9873	0.1003	0.1011	0.0418	96%
(PLS, Probit)	0.9872	0.1003	0.1011	0.0417	96%
(PLS, Logit)	0.9873	0.1002	0.1010	0.0424	96%
(PIR, SIR)	0.9865	0.1050	0.1059	0.0430	96%
(PIR, Probit)	0.9863	0.1051	0.1060	0.0429	96%
(PIR, Logit)	0.9864	0.1050	0.1059	0.0435	96%
For β_{31}:					
OLS*	1.0008	0.0455	0.0455	0.0458	96%
OLS-S	2.1159	0.1184	1.1222	0.1152	–
PILS-SDR*	1.0018	0.0724	0.0724	0.0694	95%
PILS-SDR:					
(SIR, SIR)	1.0073	0.0977	0.0980	0.1072	98%
(SIR, Probit)	1.0082	0.0973	0.0977	0.1058	98%
(SIR, Logit)	1.0083	0.0975	0.0979	0.1090	98%
(PLS, SIR)	1.0066	0.0979	0.0981	0.1047	97%
(PLS, Probit)	1.0076	0.0973	0.0976	0.1033	97%
(PLS, Logit)	1.0077	0.0974	0.0977	0.1068	97%
(PIR, SIR)	1.0072	0.0980	0.0983	0.1130	98%
(PIR, Probit)	1.0081	0.0977	0.0980	0.1121	98%
(PIR, Logit)	1.0082	0.0978	0.0982	0.1155	98%

Table D2: (continued)

Panel F: Distribution 2 (non-elliptical) - Model C: $g_{21}(z) = [2\mathbb{I}\{z > 0\} + 1]\sqrt{|z|}$

Estimator	*Mean*	*SD*	*RMSE*	*MedSE*	*CR*
Homogeneous Populations: $\left(\rho_U^A, \rho_D^A\right) = \left(\rho_U^B, \rho_D^B\right) = (0.35, 0.40)$					
For β_{21}:					
OLS*	0.9994	0.0191	0.0191	0.0195	96%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9844	0.1243	0.1253	0.0514	95%
PILS-SDR:					
(SIR, SIR)	0.9854	0.1078	0.1087	0.0577	96%
(SIR, Probit)	0.9850	0.1077	0.1087	0.0580	96%
(SIR, Logit)	0.9851	0.1076	0.1086	0.0591	96%
(PLS, SIR)	0.9833	0.1106	0.1119	0.0580	96%
(PLS, Probit)	0.9828	0.1106	0.1120	0.0584	95%
(PLS, Logit)	0.9830	0.1106	0.1119	0.0596	96%
(PIR, SIR)	0.9867	0.1024	0.1032	0.0591	96%
(PIR, Probit)	0.9863	0.1023	0.1032	0.0594	96%
(PIR, Logit)	0.9864	0.1023	0.1031	0.0604	96%
For β_{31}:					
OLS*	1.0013	0.0448	0.0448	0.0455	96%
OLS-S	1.7714	0.0937	0.7770	0.0932	–
PILS-SDR*	1.0019	0.0712	0.0713	0.0696	94%
PILS-SDR:					
(SIR, SIR)	1.0008	0.0842	0.0842	0.0898	97%
(SIR, Probit)	1.0021	0.0836	0.0836	0.0910	97%
(SIR, Logit)	1.0022	0.0837	0.0838	0.0949	97%
(PLS, SIR)	0.9996	0.0871	0.0871	0.0905	96%
(PLS, Probit)	1.0009	0.0866	0.0866	0.0919	97%
(PLS, Logit)	1.0010	0.0866	0.0866	0.0962	97%
(PIR, SIR)	1.0004	0.0839	0.0839	0.0943	97%
(PIR, Probit)	1.0018	0.0834	0.0834	0.0955	98%
(PIR, Logit)	1.0019	0.0835	0.0835	0.0998	98%
Heterogeneous Populations: $\left(\rho_U^A, \rho_D^A\right) = (0.35, 0.40)$, $\left(\rho_U^B, \rho_D^B\right) = (-0.35, -0.40)$					
For β_{21}:					
OLS*	0.9994	0.0191	0.0191	0.0195	96%
OLS-S	–	–	–	–	–
PILS-SDR*	0.9924	0.0885	0.0889	0.0514	96%
PILS-SDR:					
(SIR, SIR)	0.9762	0.1094	0.1119	0.0590	95%
(SIR, Probit)	0.9759	0.1092	0.1118	0.0594	95%
(SIR, Logit)	0.9761	0.1092	0.1118	0.0608	96%
(PLS, SIR)	0.9742	0.1151	0.1180	0.0590	94%
(PLS, Probit)	0.9739	0.1150	0.1179	0.0593	94%
(PLS, Logit)	0.9740	0.1150	0.1179	0.0608	95%
(PIR, SIR)	0.9770	0.1074	0.1098	0.0603	96%
(PIR, Probit)	0.9768	0.1072	0.1097	0.0605	96%
(PIR, Logit)	0.9769	0.1072	0.1097	0.0618	96%
For β_{31}:					
OLS*	1.0013	0.0448	0.0448	0.0455	96%
OLS-S	1.7714	0.0937	0.7770	0.0932	–
PILS-SDR*	1.0007	0.0700	0.0700	0.0695	94%
PILS-SDR:					
(SIR, SIR)	1.0047	0.0876	0.0877	0.0941	97%
(SIR, Probit)	1.0056	0.0870	0.0872	0.0955	97%
(SIR, Logit)	1.0057	0.0871	0.0873	0.1001	98%
(PLS, SIR)	1.0032	0.0916	0.0916	0.0934	95%
(PLS, Probit)	1.0040	0.0908	0.0909	0.0950	96%
(PLS, Logit)	1.0041	0.0909	0.0910	0.1000	97%
(PIR, SIR)	1.0045	0.0878	0.0880	0.0984	97%
(PIR, Probit)	1.0055	0.0873	0.0875	0.0994	97%
(PIR, Logit)	1.0055	0.0874	0.0875	0.1044	98%

Notes: *Mean* = simulation average of the parameter estimate; *SD* = simulation standard deviation of the parameter estimate; *RMSE* = root mean-squared error of the parameter estimate; *MedSE* = simulation median of the standard error; and *CR* = coverage rate against the nominal 95% confidence interval.

Supplement E: Corrigendum of Hirukawa et al. (2023)

We have found that Ω, the key component in the asymptotic variance V given in Theorem 2 of Hirukawa et al. (2023), abbreviated as "HMP23" as in the main text, is incorrect. This error affects several parts of HMP23. In this Supplement we provide the correct expression of Ω. Correspondingly, we modify the covariance estimation in Section 2.4.3 of HMP23 and the corresponding parts of Table 1 in Section 3 (simulation study) and Table 4 in Section 4 (real data analysis) of HMP23. In what follows, we use original notations from HMP23 to ensure continuity.

The correct form of Ω in Theorem 2 of HMP23 is

$$\Omega := \Omega_1 + \kappa\Omega_2 := E\left(X_i X_i^\top \epsilon_i^2\right) + \kappa E\left\{G\left(X_{3j}\right) G\left(X_{3j}\right)^\top \beta_2^\top \eta_{2j} \eta_{2j}^\top \beta_2\right\},$$

where $G\left(x_3\right) := E\left(X \mid X_3 = x_3\right) = \left(1, E\left(X_1 \mid X_3 = x_3\right)^\top, g_2\left(x_3\right)^\top, x_{3I}^\top\right)^\top$. While the formula of Ω_1 stands, that of Ω_2 is revised from what was originally presented. These two covariance matrices are outer products of two independent influence functions $\psi_{1i} := X_i \epsilon_i$ and $\psi_{2j} := G\left(X_{3j}\right) \eta_{2j}^\top \beta_2$, which reflect the sampling error and the approximation error of $\hat{g}_2(\cdot)$ to $g_2(\cdot)$, respectively. The errors arise in $\mathcal{S}_1$ and $\mathcal{S}_2$, and the subscripts "i" and "j" in the formulae of Ω_1 and Ω_2 are intended to distinguish sources of the errors. Because a minor modification of Lemma A4 in Supplement A offers the correct expression of Ω_2, we omit the proof to save space.

Now we provide a consistent estimator of Ω. It takes the general form of $\hat{\Omega} = \hat{\Omega}_1 + (n/m)\hat{\Omega}_2$, where $\hat{\Omega}_1$ has been already given in Section 2.4.3 of HMP23. The remaining task to to deliver a consistent estimator of Ω_2. Obviously, its sample analog

$$\hat{\Omega}_2 := \frac{1}{m} \sum_{j=1}^{m} \hat{G}\left(X_{3j}\right) \hat{G}\left(X_{3j}\right)^\top \hat{\beta}_{2,PILS}^\top \hat{\eta}_{2j} \hat{\eta}_{2j}^\top \hat{\beta}_{2,PILS}$$

suffices, where $\hat{G}(\cdot)$ is some consistent estimator of $G(\cdot)$, and $\hat{\eta}_{2j} := X_{2j} - \hat{g}_{2,-j}\left(X_{3j}\right)$ is the nonparametric regression residual with $\hat{g}_{2,-j}\left(X_{3j}\right)$ being the leave-one-out NW estimator with the j-th observation eliminated. To compute $\hat{G}\left(X_{3j}\right)$, we need to estimate $E\left(X_1 \mid X_3 = X_{3j}\right)$ additionally. It can be estimated by

$$\hat{E}\left(X_1 \mid X_3 = X_{3j}\right) := \frac{\sum_{i=1}^{n} X_{1i} \mathcal{W}\left(X_{3i}; X_{3j}, \mathbf{a}, \lambda\right)}{\sum_{i=1}^{n} \mathcal{W}\left(X_{3i}; X_{3j}, \mathbf{a}, \lambda\right)},$$

where $\mathbf{a} = \left(a_1, \ldots, a_{d_{3C}}\right)$ for $a \in \{h, b\}$ and $\lambda = \left(\lambda_1, \ldots, \lambda_{d_{3D}}\right)$ are sets of tuning parameters for continuous and discrete common variables, respectively. In practice, for the Epanechnikov and beta kernels, we can employ $\hat{h}_p = \hat{\sigma}_{X_p} (\log n/n)^{0.3}$ and

$\hat{b}_p = \hat{\sigma}_{U_p} (\log n/n)^{0.6}$ for $p \in \{1, \ldots, d_{3C}\}$, respectively, where $\hat{\sigma}_{X_p}$ and $\hat{\sigma}_{U_p}$ are sample standard deviations of the corresponding continuous common variable in the original scale and in the transformed scale on $[0, 1]$, respectively. For the discrete kernel, we may put $\hat{\lambda}_q = (\log n/n)^{0.6}$ for $q \in \{1, \ldots, d_{3D}\}$, regardless of whether the discrete common variable is ordered or unordered.

Finally, the corrected covariance estimation affects both the Monte Carlo results in Section 3 and the estimation results of a real data analysis in Section 4 of HMP23. More specifically, $\overline{SE}$ (simulation average of the standard error) and CR (coverage rate against the nominal 95% confidence interval) for "PILS-E" and "PILS-B" in Table 1 of HMP23 are recalculated and provided in Table E1 (corrections are underlined). In addition, standard errors under the heading "PILS" in Table 4 of HMP23 are replaced. Table E4 is the replacement. These changes are underlined. It is worth noting that the impact of revising the standard errors formula is minimal and leads to no changes of conclusions.

Table E1: Monte Carlo results (corrections are underlined.)

	$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*
Model A ($\rho_1=0.1$)										
OLS*	0.9986	0.0559	0.0559	0.0560	95%	1.0020	0.0397	0.0398	0.0397	95%
OLS-S	2.9802	0.0450	1.9807	–	–	2.9807	0.0314	1.9809	–	–
IV1-S	0.7461	1.2942	1.3189	1.1668	92%	0.9019	0.6212	0.6289	0.6171	94%
IV2-S	0.7076	1.5997	1.6262	1.3316	93%	0.8921	0.6651	0.6738	0.6457	93%
2SLS-S	1.0207	0.9247	0.9249	0.9019	88%	1.0115	0.5629	0.5630	0.5744	92%
GMM-S	1.0209	0.9251	0.9253	0.9012	88%	1.0112	0.5631	0.5632	0.5741	92%
MSOLS	1.1265	0.0624	0.1411	–	–	1.0938	0.0434	0.1034	–	–
MSII	0.9058	0.0805	0.1239	–	–	0.9467	0.0504	0.0734	–	–
MSII-FM	0.8203	0.0827	0.1978	0.1066	67%	0.9052	0.0511	0.1077	0.0629	72%
PARA	0.9990	0.0612	0.0612	0.0611	95%	1.0009	0.0426	0.0426	0.0432	96%
PILS-E	1.0288	0.0720	0.0775	0.1041	99%	1.0024	0.0462	0.0463	0.0640	99%
PILS-B	1.0013	0.0637	0.0637	0.0969	99%	0.9963	0.0437	0.0439	0.0619	99%
Model A ($\rho_1=0.4$)										
OLS*	0.9999	0.0294	0.0294	0.0293	94%	1.0012	0.0204	0.0205	0.0207	96%
OLS-S	1.4203	0.0451	0.4228	–	–	1.4204	0.0311	0.4216	–	–
IV1-S	0.9993	0.1163	0.1163	0.1165	96%	0.9970	0.0800	0.0801	0.0821	96%
IV2-S	0.9989	0.1198	0.1198	0.1202	95%	0.9965	0.0832	0.0832	0.0848	96%
2SLS-S	1.0018	0.1165	0.1165	0.1161	95%	0.9982	0.0798	0.0799	0.0820	96%
GMM-S	1.0019	0.1167	0.1168	0.1160	96%	0.9982	0.0798	0.0798	0.0819	96%
MSOLS	1.0166	0.0313	0.0354	–	–	1.0152	0.0219	0.0266	–	–
MSII	0.9952	0.0317	0.0321	–	–	0.9987	0.0221	0.0221	–	–
MSII-FM	0.9861	0.0317	0.0347	0.0308	92%	0.9938	0.0221	0.0229	0.0216	94%
PARA	0.9998	0.0310	0.0310	0.0307	95%	1.0010	0.0214	0.0214	0.0217	96%
PILS-E	1.0094	0.0354	0.0367	0.0357	93%	1.0035	0.0236	0.0239	0.0238	95%
PILS-B	1.0013	0.0318	0.0318	0.0326	96%	1.0012	0.0219	0.0219	0.0226	96%
Model B ($\rho_1=0.1$)										
OLS*	0.9986	0.0602	0.0602	0.0613	95%	0.9996	0.0447	0.0447	0.0433	95%
OLS-S	1.4334	0.0288	0.4344	–	–	1.4343	0.0199	0.4347	–	–
IV1-S	0.9359	0.4259	0.4306	0.4207	97%	0.9820	0.2432	0.2439	0.2432	97%
IV2-S	0.9204	0.5361	0.5420	0.4801	96%	0.9794	0.2593	0.2602	0.2537	97%
2SLS-S	0.9978	0.3717	0.3717	0.3623	95%	1.0067	0.2299	0.2300	0.2330	96%
GMM-S	0.9974	0.3721	0.3721	0.3619	94%	1.0064	0.2299	0.2300	0.2329	96%
MSOLS	1.0907	0.0598	0.1087	–	–	1.0663	0.0440	0.0796	–	–
MSII	0.8611	0.1102	0.1773	–	–	0.9222	0.0616	0.0992	–	–
MSII-FM	0.8204	0.1185	0.2152	0.1309	81%	0.9079	0.0630	0.1116	0.0676	77%
PARA	1.4632	2.8122	2.8501	–	–	1.5260	4.1731	4.2061	–	–
PILS-E	1.0087	0.0702	0.0707	0.0885	98%	0.9926	0.0487	0.0492	0.0575	98%
PILS-B	0.9607	0.0727	0.0827	0.1033	99%	0.9636	0.0507	0.0624	0.0626	97%
Model B ($\rho_1=0.4$)										
OLS*	0.9990	0.0371	0.0371	0.0373	96%	1.0005	0.0271	0.0271	0.0263	94%
OLS-S	1.1077	0.0283	0.1113	–	–	1.1084	0.0190	0.1101	–	–
IV1-S	0.9979	0.0687	0.0687	0.0703	96%	1.0004	0.0496	0.0496	0.0496	95%
IV2-S	0.9971	0.0709	0.0710	0.0726	95%	1.0002	0.0511	0.0511	0.0512	96%
2SLS-S	0.9987	0.0689	0.0689	0.0702	96%	1.0008	0.0494	0.0494	0.0495	95%
GMM-S	0.9987	0.0690	0.0691	0.0701	95%	1.0008	0.0494	0.0494	0.0495	95%
MSOLS	1.0404	0.0370	0.0548	–	–	1.0378	0.0265	0.0462	–	–
MSII	0.9854	0.0538	0.0557	–	–	0.9917	0.0362	0.0371	–	–
MSII-FM	0.9809	0.0553	0.0585	0.0534	93%	0.9897	0.0366	0.0380	0.0355	94%
PARA	1.1324	0.9097	0.9192	–	–	1.1148	2.0242	2.0275	–	–
PILS-E	1.0192	0.0414	0.0456	0.0435	92%	1.0110	0.0298	0.0318	0.0305	94%
PILS-B	0.9992	0.0417	0.0418	0.0447	96%	0.9992	0.0307	0.0307	0.0307	96%

Table E4: Estimation results (corrections are underlined.)

	(1) OLS*	(2) OLS-S	(3) IV-S	(4) MSOLS	(5) MSII-FM	(6) PARA	(7) PILS
Education	0.0635	0.0718	0.0953	0.0727	0.0685	−0.2850	0.0568
	(0.0056)	(0.0053)	(0.0159)	(0.0059)	(0.0080)	(0.2281)	(0.0093)
Experience	0.0809	0.0818	0.0830	0.0826	0.0766	0.1170	0.0818
	(0.0043)	(0.0043)	(0.0045)	(0.0049)	(0.0065)	(0.0095)	(0.0043)
$Experience^2$	−0.0017	−0.0017	−0.0017	−0.0017	−0.0016	−0.0017	−0.0017
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Ability	0.0313	−	−	−0.0012	0.0029	0.2444	0.0094
	(0.0078)	(−)	(−)	(0.0027)	(0.0081)	(0.1514)	(0.0048)
Married	0.3717	0.3793	0.3844	0.3799	0.3777	0.2226	0.3958
	(0.0539)	(0.0536)	(0.0536)	(0.0535)	(0.0535)	(0.1319)	(0.0558)
Black	−0.1302	−0.1741	−0.1249	−0.1849	−0.1504	−0.1819	−0.1776
	(0.0323)	(0.0316)	(0.0426)	(0.0393)	(0.0806)	(0.1146)	(0.0322)
South	−0.0921	−0.0983	−0.0814	−0.0979	−0.0989	−0.1372	−0.1042
	(0.0288)	(0.0287)	(0.0314)	(0.0286)	(0.0289)	(0.1135)	(0.0292)
Urban	0.1363	0.1499	0.1278	0.1538	0.1404	0.0138	0.1557
	(0.0282)	(0.0284)	(0.0332)	(0.0297)	(0.0390)	(0.1273)	(0.0288)
Data combination?	No	No	No	Yes	Yes	Yes	Yes
Sample size: n	2430	2430	2430	2430	2430	2430	2430
m	−	−	−	1102	1102	1102	1102

References

[1] Abadie, A., and G. W. Imbens (2016): "Matching on the Estimated Propensity Score," *Econometrica*, 84, 781 - 807.

[2] Efron, B. (1992): "Jackknife-After-Bootstrap Standard Errors and Influence Functions," *Journal of the Royal Statistical Society, Series B*, 54, 83 - 127.

[3] Hansen, B. E. (2008): "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory*, 24, 726 - 748.

[4] Hirukawa, M., D. Liu, and A. Prokhorov (2021): "msreg: A Stata Command for Consistent Estimation of Linear Regression Models Using Matched Data," *Stata Journal*, 21, 123 - 140.

[5] Hirukawa, M., and A. Prokhorov (2018): "Consistent Estimation of Linear Regression Models Using Matched Data," *Journal of Econometrics*, 203, 344 - 358.

[6] Hirukawa, M., I. Murtazashvili, and A. Prokhorov (2023): "Yet Another Look at the Omitted Variable Bias," *Econometric Reviews*, 42, 1 - 27.

[7] Horowitz, J. L., and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599 - 631.

[8] Kulasekera, K. B., and W. Lin (2010): "Error Variance Estimation for the Single-Index Model," *Australian & New Zealand Journal of Statistics*, 52, 201 - 219.

[9] Lehmann, E. L. (1999): *Elements of Large-Sample Theory*. New York: Springer-Verlag.

[10] Neter, J., E. S. Maynes, and R. Ramanathan (1965): "The Effect of Mismatching on the Measurement of Response Error," *Journal of the American Statistical Association*, 60, 1005 - 1027.

[11] Shao, J. (1989): "The Efficiency and Consistency of Approximations to the Jackknife Variance Estimators," *Journal of the American Statistical Association*, 84, 114 - 119.

[12] Ullah, A. (1988): "Nonparametric Estimation and Hypothesis Testing in Econometric Models," *Empirical Economics*, 13, 223 - 249.

[13] White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1 - 25.

[14] Yang, S., and J. K. Kim (2020): "Asymptotic Theory and Inference of Predictive Mean Matching Imputation Using a Superpopulation Model Framework," *Scandinavian Journal of Statistics*, 47, 839 - 861.