Supplement to "Inference on Semiparametric Multinomial Response Models"

Shakeeb Khan[*] Boston College

Fu Ouyang[†] University of Queensland

Elie Tamer[‡] Harvard University

December 4, 2020

These appendixes contain three sets of results. Appendix A proves Theorems 2.1 and 2.2 for the cross-sectional MRC estimator. Appendix B derives the asymptotic properties of the MS estimators proposed in for the static and dynamic panel data models. Appendix C collects additional Monte Carlo results.

A Cross-Sectional Estimator

Proof of Theorem 2.1. It suffices to show the identification of β_0 based on (2.3) as the same arguments can be applied to similar identification inequalities. Denote $\Omega_{im} = \{x_{i2} = x_{m2}\}$. By Assumptions CS1 and CS2, the monotonic relation (2.3) implies that β_0 maximizes

$$G_1(b) \equiv E\{E[y_{i1} - y_{m1}|x_i, x_m, \Omega_{im}] \cdot sgn(x'_{im1}b)|\Omega_{im}\}$$

for each pair of (i, m). To show that β_0 attains a unique maximum, let $b \in \mathcal{B}$ such that $G_1(b) = G_1(\beta_0)$. We assume $\beta_0^{(1)} = b^{(1)} = 1$ w.l.o.g. (the case $\beta_0^{(1)} = b^{(1)} = -1$ is symmetric). We want to show that $\tilde{b} = \tilde{\beta}_0$ must hold. To see this, first note that if $P[(\tilde{x}'_{im1}\tilde{\beta}_0 < -x^{(1)}_{im1} < \tilde{x}'_{im1}\tilde{b}) \cup (\tilde{x}'_{im1}\tilde{b} < -x^{(1)}_{im1} < \tilde{x}'_{im1}\tilde{\beta}_0)|\Omega_{im}] > 0$, β_0 and b yield different values of the $sgn(\cdot)$ function in $G_1(\cdot)$ with strictly positive probability, and thus $G_1(b) < G_1(\beta_0)$. This implies that for all b satisfying $G_1(b) = G_1(\beta_0)$, $P[(\tilde{x}'_{im1}\tilde{\beta}_0 < -x^{(1)}_{im1} < \tilde{x}'_{im1}\tilde{b}) \cup (\tilde{x}'_{im1}\tilde{b} < -x^{(1)}_{im1} < \tilde{x}'_{im1}\tilde{\beta}_0)|\Omega_{im}] = 0$ must hold, which is equivalent to $P(\tilde{x}'_{im1}\tilde{\beta}_0 = \tilde{x}'_{im1}\tilde{b}|\Omega_{im}) = 1$ under Assumption CS3(i). Then the desired result follows from Assumption CS3(ii). □

[*]Email: shakeeb.khan@bc.edu.

[†]Email: f.ouyang@uq.edu.au.

[‡]Email: elietamer@fas.harvard.edu.

We next derive the asymptotic properties of $\hat{\beta}$ summarized in Theorem 2.2. We first prove Lemmas A.1 and A.2 to establish the consistency of $\hat{\beta}$. Then we show Lemmas A.3–A.5 to establish the asymptotic normality of $\hat{\beta}$. For ease of illustration, with a bit abuse of notation, we work with objective function

$$G_{1n}^K(b) \equiv \frac{1}{n(n-1)h_n^p} \sum_{i \neq m} K_{h_n}(x_{im2}) q_{im}(b)$$

and $G_1(b) \equiv f_2(0) E[q_{im}(b)|\Omega_{im}]$, where $q_{im}(b) \equiv y_{im1} \cdot [sgn(x'_{im1}b) - sgn(x'_{im1}\beta_0)]$.[1] Accordingly, we will also subtract the term $sgn(x'_{im1}\beta_0)$ for $S_{im}(b)$ and $\tau_i(b)$ in this appendix.

Lemma A.1. *If Assumptions CS1–CS6, CS8, and CS9 hold,* $\sup_{b \in \mathcal{B}} |G_{1n}^K(b) - G_1(b)| = o_p(1)$.

Proof of Lemma A.1. Note that Lemma A.1 follows from $\sup_{b \in \mathcal{B}} |G_{1n}^K(b) - E[G_{1n}^K(b)]| = o_p(1)$ and $\sup_{b \in \mathcal{B}} |E[G_{1n}^K(b)] - G_1(b)| = o(1)$.

Define $\mathcal{F}_n = \{K_{h_n}(x_{im2}) q_{im}(b) | b \in \mathcal{B}\}$ with $h_n > 0$ and $h_n \to 0$. $\mathcal{F}_n$ is a sub-class of the fixed class $\mathcal{F} \equiv \{K(x_{im2}/h) q_{im}(b) | b \in \mathcal{B}, h > 0\} = \mathcal{F}_h \mathcal{F}_b$, with $\mathcal{F}_h \equiv \{K(x_{im2}/h) | h > 0\}$, which is Euclidean for the constant envelope $\sup_{v \in \mathbb{R}^p} |K(v)|$ by Lemma 22 in Nolan and Pollard (1987) and $\mathcal{F}_b \equiv \{q_{im}(b) | b \in \mathcal{B}\}$. Noticing that $q_{im}(b)$ is uniformly bounded by 2, Example 2.11 and Lemma 2.15 in Pakes and Pollard (1989) then implies that $\mathcal{F}_b$ is Euclidean for the constant envelope 2. Putting all these results together, we conclude using Lemma 2.14 in Pakes and Pollard (1989) that $\mathcal{F}$ is Euclidean for the constant envelope $2 \sup_{v \in \mathbb{R}^p} |K(v)|$. Applying Corollary 4 in Sherman (1994a), we obtain that $\sup_{b \in \mathcal{B}} |G_{1n}^K(b) - E[G_{1n}^K(b)]| = O_p(n^{-1}/h_n^p) = o_p(1)$ by Assumption CS9.

It remains to show that $\sup_{b \in \mathcal{B}} |E[G_{1n}^K(b)] - G_1(b)| = o(1)$. Letting $\varphi(\cdot) \equiv f_2(\cdot) E[q_{im}(b) | x_{im2} = \cdot]$, we can write by Assumptions CS5, CS6, CS8, and CS9 that

$$\begin{aligned}
&\sup_{b \in \mathcal{B}} |E[G_{1n}^K(b)] - G_1(b)| = \sup_{b \in \mathcal{B}} |h_n^{-p} E[K(x_{im2}/h_n) q_{im}(b)] - \varphi(0)| \\
=& \sup_{b \in \mathcal{B}} |\int h_n^{-p} K(v/h_n) \varphi(v) dv - \varphi(0)| = \sup_{b \in \mathcal{B}} |\int h_n^{-p} K(v/h_n)[\varphi(0) + v' \nabla_1 \varphi(\bar{v})] dv - \varphi(0)| \\
=& \sup_{b \in \mathcal{B}} |\int K(u)[\varphi(0) + h_n u' \nabla_1 \varphi(\bar{u}_n)] du - \varphi(0)| = h_n \cdot \sup_{b \in \mathcal{B}} |\int K(u) u' \nabla_1 \varphi(\bar{u}_n) du| \\
\leq& h_n \cdot \int |K(u)||u|_1 \sup_{b \in \mathcal{B}} |\nabla_1 \varphi(\bar{u}_n)|_1 du = O(h_n) = o(1),
\end{aligned}$$

where the third equality applies a mean-value expansion and the fourth equality uses a change of variables $u = v/h_n$. Combining all these results completes the proof. □

Lemma A.2. *If Assumptions CS1–CS6, CS8, and CS9 hold,* $\hat{\beta} \xrightarrow{p} \beta_0$.

Proof of Lemma A.2. The proof proceeds by verifying the four sufficient conditions for Theorem

[1] Here we subtract the term $y_{im1} \cdot sgn(x'_{im1}\beta_0)$ from objective function (2.6), analogous to Sherman (1993). Doing this does not affect the value of the estimator, and will facilitate the proofs that follow.

2.1 in Newey and McFadden (1994): (C1) $\mathcal{B}$ is a compact set, (C2) $\sup_{b\in\mathcal{B}} |G_{1n}^K(b) - G_1(b)| = o_p(1)$, (C3) $G_1(b)$ is continuous in b, and (C4) $G_1(b)$ is uniquely maximized at β_0.

Condition (C1) is satisfied by Assumption CS4. Lemma A.1 secures condition (C2). The identification condition (C4) is essentially verified in the proof of Theorem 2.1, given that $f_2(0) > 0$ by Assumption CS5.

It remains to verify the continuity of $G_1(b)$ in b. Assuming $b^{(1)} = 1$ w.l.o.g., $E[q_{im}(b)|\Omega_{im}]$ can be expressed as the sum of terms like

$$P(y_{im1} = d, x_{im1}^{(1)} > -\tilde{x}_{im1}'\tilde{b}|\Omega_{im}) = \int\int_{-\tilde{x}_{im1}'\tilde{b}}^{\infty} P(y_{im1} = d|x_{im1}, \Omega_{im}) f_{x_{im1}^{(1)}|\tilde{x}_{im1},\Omega_{im}}(x) dx dF_{\tilde{x}_{im1}|\Omega_{im}}$$

for some $d \in \{-1, 0, 1\}$. Then $G_1(b)$ is continuous in b if $f_{x_{im1}^{(1)}|\tilde{x}_{im1},\Omega_{im}}(\cdot)$ does not have any mass points, which is guaranteed by Assumption CS3. This completes the proof. □

With consistency proved, the next steps turn to establishing asymptotic normality of $\hat{\beta}$, in which the derivation can be within a shrinking neighborhood of β_0, $\mathcal{B}_n \equiv \{b \in \mathcal{B} | \|b - \beta_0\| \leq \delta_n\}$ with $\delta_n = O(n^{-\delta})$ for some $0 < \delta \leq 1/2$. Our goal is to apply Theorem 2 of Sherman (1994a), of which a sufficient condition is that uniformly over $O_p(n^{-1/2})$ neighborhoods of β_0,

$$G_{1n}^K(b) = \frac{1}{2}(b - \beta_0)'V(b - \beta_0) + \frac{1}{\sqrt{n}}(b - \beta_0)'W_n + o_p(n^{-1}), \tag{A.1}$$

where V is a negative definite matrix and W_n is asymptotically normal, with mean zero and variance Λ. The verification of (A.1) has two logical steps. First, we show that

$$G_{1n}^K(b) = \frac{1}{2}(b - \beta_0)'V(b - \beta_0) + \frac{1}{\sqrt{n}}(b - \beta_0)'W_n + o_p(\|b - \beta_0\|^2) + O_p(\varepsilon_n) \tag{A.2}$$

uniformly over $\mathcal{B}_n$. Theorem 1 of Sherman (1994b) then implies that $\hat{\beta} - \beta_0 = O_p(\sqrt{\varepsilon_n} \vee n^{-1/2})$. We then show that the $O_p(\varepsilon_n)$ term in (A.2) is of order $o_p(n^{-1})$ uniformly over $\mathcal{B}_n$. Applying Theorem 1 of Sherman (1994b) once again gives $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$, from which expression (A.1) follows.

The remainder of this section is organized as follows. We first work with the U-statistic decomposition for $G_{1n}^K(b)$ (see e.g., Sherman (1993) and Serfling (2009)),

$$G_{1n}^K(b) = E[G_{1n}^K(b)] + \left(\frac{2}{n}\sum_i E[G_{1n}^K(b)|x_i] - 2E[G_{1n}^K(b)]\right) + \rho_n(b). \tag{A.3}$$

Lemmas A.3–A.5 establish asymptotic properties of the three terms in (A.3), respectively. Combine results in these lemmas to get (A.1). Then we invoke Theorem 2 of Sherman (1994a) to prove the asymptotic normality of $\hat{\beta}$.

Lemma A.3. *If Assumptions CS1–CS9 hold, then uniformly over $\mathcal{B}_n$, we have*

$$E[G_{1n}^K(b)] = \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + o(\|b-\beta_0\|^2),$$

where $V \equiv E[\nabla_2\tau_i(\beta_0)]$.

Proof of Lemma A.3. First, we express $E[G_{1n}^K(b)]$ as the integral

$$\begin{aligned}
&\int h_n^{-p}K_{h_n}(x_{im2})B(x_{i1},x_{m1},x_{i2},x_{m2})S_{im}(b)dF_x(x_{i1},x_{i2})dF_x(x_{m1},x_{m2}) \qquad \text{(A.4)}\\
=&\int K(u_{im})B(x_{i1},x_{m1},x_{m2}+u_{im}h_n,x_{m2})S_{im}(b)f_x(x_{i1},x_{m2}+u_{im}h_n)dx_{i1}du_{im}dF_x(x_{m1},x_{m2}),
\end{aligned}$$

where we apply a change of variables $u_{im} = x_{im2}/h_n$ to obtain the equality.

Take the κ^{th}-order Taylor expansion inside the integral around x_{m2} to obtain the lead term

$$\begin{aligned}
&\int K(u_{im})B(x_{i1},x_{m1},x_{m2},x_{m2})S_{im}(b)f_x(x_{i1},x_{m2})dx_{i1}du_{im}dF_x(x_{m1},x_{m2})\\
=&\int B(x_{i1},x_{m1},x_{m2},x_{m2})S_{im}(b)f_x(x_{i1},x_{m2})dx_{i1}dF_x(x_{m1},x_{m2}) = E[\tau_m(b)], \qquad \text{(A.5)}
\end{aligned}$$

where the first equality follows by $\int K(u)du = 1$. All remaining terms are zero except the last one which is of order $O(h_n^\kappa\delta_n) = o(\|b-\beta_0\|^2)$ by Assumptions CS7–CS9.

Note that a second-order Taylor expansion around β_0 gives

$$\begin{aligned}
&\tau_m(b)-\tau_m(\beta_0) = (b-\beta_0)'\nabla_1\tau_m(\beta_0) + \frac{1}{2}(b-\beta_0)'\nabla_2\tau_m(\bar{b})(b-\beta_0)\\
=&(b-\beta_0)'\nabla_1\tau_m(\beta_0) + \frac{1}{2}(b-\beta_0)'\nabla_2\tau_m(\beta_0)(b-\beta_0) + \frac{1}{2}(b-\beta_0)'\left[\nabla_2\tau_m(\bar{b})-\nabla_2\tau_m(\beta_0)\right](b-\beta_0),
\end{aligned}$$

and hence by $\tau_m(\beta_0)=0$ and Assumption CS7,

$$\begin{aligned}
E[\tau_m(b)] =&(b-\beta_0)'E[\nabla_1\tau_m(\beta_0)] + \frac{1}{2}(b-\beta_0)'E[\nabla_2\tau_m(\beta_0)](b-\beta_0) + o(\|b-\beta_0\|^2)\\
=&\frac{1}{2}(b-\beta_0)'V(b-\beta_0) + o(\|b-\beta_0\|^2), \qquad \text{(A.6)}
\end{aligned}$$

where the second equality uses the fact that $E[\nabla_1\tau_m(\beta_0)] = 0$ as $E[\tau_m(b)]$ is maximized at β_0. Then, applying (A.4), (A.5), and (A.6) proves the lemma. □

Lemma A.4. *If Assumptions CS1– CS9, then uniformly over $\mathcal{B}_n$, we have*

$$\frac{2}{n}\sum_i E[G_{1n}^K(b)|x_i] - 2E[G_{1n}^K(b)] = \frac{1}{\sqrt{n}}(b-\beta_0)'W_n + o_p(\|b-\beta_0\|^2),$$

where $W_n \equiv n^{-1/2}\sum_i 2\nabla_1\tau_i(\beta_0)$.

Proof of Lemma A.4. We establish a representation for $E[G^K_{1n}(b)|x_m]$ using the same arguments as in the proof of Lemma A.3, but no longer integrating over x_m. Specifically, a change of variables $u_{im} = x_{im2}/h_n$ gives

$$
\begin{aligned}
E[G^K_{1n}(b)|x_m] &= \int h_n^{-p} K_{h_n}(x_{im2}) B(x_{i1}, x_{m1}, x_{i2}, x_{m2}) S_{im}(b) dF_x(x_{i1}, x_{i2}) \\
&= \int K(u_{im}) B(x_{i1}, x_{m1}, x_{m2} + u_{im}h_n, x_{m2}) S_{im}(b) f_x(x_{i1}, x_{m2} + u_{im}h_n) dx_{i1} du_{im}.
\end{aligned}
$$

The lead term of the κ^{th}-order Taylor expansion inside the integral around x_{m2} is

$$
\begin{aligned}
&\int K(u_{im}) B(x_{i1}, x_{m1}, x_{m2}, x_{m2}) S_{im}(b) f_x(x_{i1}, x_{m2}) dx_{i1} du_{im} \\
=& \int B(x_{i1}, x_{m1}, x_{m2}, x_{m2}) S_{im}(b) f_x(x_{i1}, x_{m2}) dx_{i1} = \tau_m(b)
\end{aligned}
$$

and the sample average of the bias term is of order $o_p(\|b - \beta_0\|^2)$.

Then, applying Lemma A.3 and Assumption CS7, we write

$$
\begin{aligned}
&\frac{2}{n}\sum_m E[G^K_{1n}(b)|z_m] - 2E[G^K_{1n}(b)] \\
=&\frac{2}{n}\sum_m \tau_m(b) - (b-\beta_0)'V(b-\beta_0) + o_p(\|b-\beta_0\|^2) \\
=&\frac{2}{n}\sum_m (b-\beta_0)'\nabla_1\tau_m(\beta_0) + \frac{1}{n}\sum_m (b-\beta_0)'\left(\nabla_2\tau_m(\beta_0) - V\right)(b-\beta_0) + o_p(\|b-\beta_0\|^2).
\end{aligned}
$$

Then the desired result follows as $n^{-1}\sum_m \nabla_2\tau_m(\beta_0) - V = o_p(1)$ by the SLLN. □

Lemma A.5. *If Assumptions CS1–CS9 hold, then $\rho_n(b) = O_p(n^{-1}h_n^{-p})$ uniformly over $\mathcal{B}_n$.*

Proof of Lemma A.5. By construction, we can write

$$
\rho_n(b) = \frac{1}{n(n-1)}\sum_{i\neq m} \rho_{im}(b),
$$

where $\rho_{im}(b) \equiv (K_{h_n}(x_{im})q_{im}(b) - 2n^{-1}\sum_i E[K_{h_n}(x_{im})q_{im}(b)|x_i] + E[K_{h_n}(x_{im})q_{im}(b)])/h_n^p$.

Note that $\rho_{im}(\beta_0) = 0$ and $|\rho_{im}(b)|$ is bounded by a multiple of M/h_n^p, where M is a positive constant. Define $\mathcal{F}_n^* = \{\rho^*_{im}(b)|b \in \mathcal{B}\}$ with $\rho^*_{im}(b) \equiv h_n^p\rho_{im}(b)/M$. Then the Euclidean properties of the (P-degenerate) class of functions $\mathcal{F}_n^*$ are deduced using similar arguments for proving Lemma A.1 in combination with Corollaries 17 and 21 in Nolan and Pollard (1987). As $\sup_{b\in\mathcal{B}_n} E[\rho^*_{im}(b)^2] =$

$O(1)$ by Assumption CS8, applying Theorem 3 of Sherman (1994a) gives

$$\frac{1}{n(n-1)}\sum_{i\neq m}\rho^*_{im}(b) = O_p(n^{-1}),$$

and hence $\rho_n(b) = O_p(n^{-1}h_n^{-p})$. □

Proof of Theorem 2.2. We have shown part (i) of Theorem 2.2 (consistency of $\hat{\beta}$) in Lemma A.2. Here we move on to prove part (ii). Putting results in Lemmas A.3–A.5 together, we write

$$G^K_{1n}(b) = \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + \frac{1}{\sqrt{n}}(b-\beta_0)'W_n + o_p(\|b-\beta_0\|^2) + O_p(\varepsilon_n) \tag{A.7}$$

where $\varepsilon_n = n^{-1}h_n^{-p}$. Theorem 1 of Sherman (1994b) then implies that $\hat{\beta}-\beta_0 = O_p(\sqrt{\varepsilon_n})$.

Next, take $\delta_n = O(\sqrt{\varepsilon_n})$ and $\mathcal{B}_n = \{b \in \mathcal{B} \| \|b-\beta_0\| \leq \delta_n\}$. We repeat the proof for Lemma A.5 and deduce from a Taylor expansion around β_0 that $E[\sup_{b\in\mathcal{B}_n}\rho^*_{im}(b)^2] = O(\delta_n^2 h_n^p)$. Apply Theorem 3 of Sherman (1994b) to see that uniformly over $\mathcal{B}_n$, $\rho_n(b) = O_p(n^{-1}h_n^{(\alpha/2-1)p}\delta_n^\alpha)$ for $0<\alpha<1$. Then we have

$$\rho_n(b) = O_p(n^{-1}h_n^{(\alpha/2-1)p}\delta_n^\alpha) = O_p(n^{-1})O_p(n^{-\alpha/2}h_n^{-p}) = o_p(n^{-1})$$

by invoking Assumption CS9 and choosing α sufficiently close to 1. This result in turn implies that the $O_p(\varepsilon_n)$ term in (A.7) is of order $o_p(n^{-1})$, and hence $\hat{\beta}-\beta_0 = O_p(n^{-1/2})$ by applying Theorem 1 of Sherman (1994b) once again.

Now (A.7) can be expressed as

$$G^K_{1n}(b) = \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + \frac{1}{\sqrt{n}}(b-\beta_0)'W_n + o_p(n^{-1}).$$

Denote $\mathit{\Delta}_i = 2\nabla_1\tau_i(\beta_0)$. Note that $E[\mathit{\Delta}_i] = 0$ as $E[\nabla_1\tau_i(\beta_0)] = 0$. We deduce from Assumption CS7 and Lindeberg-Lévy CLT that $W_n \xrightarrow{d} N(0,\Lambda)$ where $\Lambda = E[\mathit{\Delta}_i\mathit{\Delta}_i']$. The asymptotic normality of $\hat{\beta}$ then follows from Theorem 2 of Sherman (1994a), i.e., $\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0, V^{-1}\Lambda V^{-1})$. □

B Static and Dynamic Panel Data Estimators

B.1 Static Panel Data Estimator

Throughout this section, we assume w.l.o.g. $\beta_0^{(1)} > 0$. The case $\beta_0^{(1)} < 0$ is symmetric. To lighten the notation, we suppress the subscript i whenever it is clear from the context that all variables are for each individual. With a bit abuse of notation, we define sample objective function

$$G_n^{SP,K}(b) = \frac{1}{nh_n^p} \sum_i K_{h_n}(z_{i1}) z_{i2} \cdot sgn(z_{i3}' b)$$

and population objective function $G^{SP}(b) = f_{z_1}(0) E[\rho(b)|z_1 = 0]$.

Let $\Omega \equiv \{z_1 = 0\}$. The following lemma establishes the identification of β_0.

Lemma B.1. *If Assumptions SP1–SP6, $G^{SP}(\beta_0) > G^{SP}(b)$ for all $b \in \mathcal{B} \setminus \{\beta_0\}$.*

Proof of Lemma B.1. Denote $\mathcal{Z}_b = \{z_3 : sgn(z_3' b) \neq sgn(z_3' \beta_0)\}$ for all $b \in \mathcal{B} \setminus \{\beta_0\}$. We write

$$\begin{aligned}
&G^{SP}(\beta_0) - G^{SP}(b) \\
=& f_{z_1}(0) E[z_2(sgn(z_3'\beta_0) - sgn(z_3' b))|\Omega] = 2 f_{z_1}(0) \int_{\mathcal{Z}_b} sgn(z_3'\beta_0) E[z_2|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_b} sgn(z_3'\beta_0) E[E[z_2|x, \alpha, \Omega]|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_b} E[sgn(z_3'\beta_0)(P(y_{11} = 1|x, \alpha, \Omega) - P(y_{12} = 1|x, \alpha, \Omega))|z_3, \Omega] dF_{z_3|\Omega}.
\end{aligned}$$

Under Assumption SP6, $f_{z_1}(0) > 0$. Furthermore, by definition,

$$P(y_{11} = 1|x, \alpha, \Omega) = P(x_{11}'\beta_0 + \alpha_1 - \epsilon_{11} > \max\{0, x_{21}'\beta_0 + \alpha_2 - \epsilon_{21}\}|x, \alpha, \Omega)$$

and

$$P(y_{12} = 1|x, \alpha, \Omega) = P(x_{12}'\beta_0 + \alpha_1 - \epsilon_{12} > \max\{0, x_{22}'\beta_0 + \alpha_2 - \epsilon_{22}\}|x, \alpha, \Omega).$$

Then Assumption SP3 implies that $sgn(P(y_{11} = 1|x, \alpha, \Omega) - P(y_{12} = 1|x, \alpha, \Omega)) = sgn(z_3'\beta_0)$. Thus, $P(y_{11} = 1|x, \alpha, \Omega) = P(y_{12} = 1|x, \alpha, \Omega)$ if and only if $z_3'\beta_0 = 0$, which is an event having zero probability measure under Assumption SP4. Then,

$$\begin{aligned}
&E[sgn(z_3'\beta_0)(P(y_{11} = 1|x, \alpha, \Omega) - P(y_{12} = 1|x, \alpha, \Omega))|z_3, \Omega] \\
=& E[|sgn(z_3'\beta_0)(P(y_{11} = 1|x, \alpha, \Omega) - P(y_{12} = 1|x, \alpha, \Omega))||z_3, \Omega] \\
=& E[|P(y_{11} = 1|x, \alpha, \Omega) - P(y_{12} = 1|x, \alpha, \Omega)||z_3, \Omega] > 0.
\end{aligned}$$

Therefore, $G^{SP}(\beta_0) - G^{SP}(b) > 0$ if and only if $P(\mathcal{Z}_b|\Omega) > 0$.

By definition, $P(\mathcal{Z}_b|\Omega) = P(x_{1(12)}^{(1)} b^{(1)} + \tilde{x}_{1(12)}' \tilde{b} > 0 > x_{1(12)}^{(1)} \beta_0^{(1)} + \tilde{x}_{1(12)}' \tilde{\beta}_0) + P(x_{1(12)}^{(1)} \beta_0^{(1)} + \tilde{x}_{1(12)}' \tilde{\beta}_0 > 0 > x_{1(12)}^{(1)} b^{(1)} + \tilde{x}_{1(12)}' \tilde{b})$. It follows by Assumption SP4 that $P(\mathcal{Z}_b|\Omega) > 0$ for all $b^{(1)} < 0$. For the case $b^{(1)} > 0$, we write

$$P(\mathcal{Z}_b|\Omega) = P(x_{1(12)}^{(1)} \in \mathcal{I} | \tilde{x}_{1(12)}' \tilde{b}/b^{(1)} \neq \tilde{x}_{1(12)}' \tilde{\beta}_0/\beta_0^{(1)}, \Omega) P(\tilde{x}_{1(12)}' \tilde{b}/b^{(1)} \neq \tilde{x}_{1(12)}' \tilde{\beta}_0/\beta_0^{(1)}|\Omega),$$

where $\mathcal{I} = \{-\tilde{x}'_{1(12)}\tilde{b}/b^{(1)} < x^{(1)}_{1(12)} < -\tilde{x}'_{1(12)}\tilde{\beta}_0/\beta^{(1)}_0\} \cup \{-\tilde{x}'_{1(12)}\tilde{\beta}_0/\beta^{(1)}_0 < x^{(1)}_{1(12)} < -\tilde{x}'_{1(12)}\tilde{b}/b^{(1)}\}$. It follows by Assumption SP4 that $P(\mathcal{Z}_b|\Omega) > 0$ if and only if $P(\tilde{x}'_{1(12)}\tilde{b}/b^{(1)} \neq \tilde{x}'_{1(12)}\tilde{\beta}_0/\beta^{(1)}_0|\Omega) > 0$.

Note that $\tilde{b}/b^{(1)} \neq \tilde{\beta}_0/\beta^{(1)}_0$ must hold, for otherwise we have $\beta^{(1)}_0 b = b^{(1)}\beta_0$, which in turn implies $b = \beta_0$ (as $\|b\| = \|\beta_0\|$ implies $b^{(1)} = \beta^{(1)}_0$). This completes the proof as by Assumption SP5, $P(\tilde{z}'_3\tilde{b}/b^{(1)} = \tilde{z}'_3\tilde{\beta}_0/\beta^{(1)}_0|\Omega) < 1$ holds true whenever $\tilde{b}/b^{(1)} \neq \tilde{\beta}_0/\beta^{(1)}_0$. □

Proof of Theorem 3.1. The proof proceeds by verifying the four sufficient conditions for applying Theorem 2.1 in Newey and McFadden (1994): (C1) $\mathcal{B}$ is a compact set, (C2) $\sup_{b\in\mathcal{B}} |G^{SP,K}_n(b) - G^{SP}(b)| = o_p(1)$, (C3) $G^{SP}(b)$ is continuous in b, and (C4) $G^{SP}(b)$ is uniquely maximized at β_0.

The compactness of $\mathcal{B}$ is satisfied by construction. Lemma B.1 above has shown that the identification condition in (C4) holds. Next, the continuity of $G^{SP}(b)$ is a result from Assumption SP4. To see this, first note that $G^{SP}(b)$ can be expressed as the sum of functions with respect to b of the following form: For some $d \in \{-1, 0, 1\}$,

$$\begin{aligned}
&P(y_{11} - y_{12} = d, x^{(1)}_{1(12)}b^{(1)} + \tilde{x}'_{1(12)}\tilde{b} > 0|\Omega) \\
&= \int\int_{-\tilde{x}'_{1(12)}\tilde{b}/b^{(1)}}^{\infty} P(y_{11} - y_{12} = d|x_{1(12)}, \Omega) f_{x^{(1)}_{1(12)}|\tilde{x}_{1(12)},\Omega}(x)dxdF_{\tilde{x}_{1(12)}|\Omega}.
\end{aligned}$$

Then, $G^{SP}(b)$ is continuous if $f_{x^{(1)}_{1(12)}|\tilde{x}_{1(12)},\Omega}(\cdot)$ does not have any mass points, which is guaranteed by Assumption SP4.

The remaining task is to verify the uniform convergence condition (C3), i.e., $\sup_{b\in\mathcal{B}} |G^{SP,K}_n(b) - G^{SP}(b)| = o_p(1)$. It suffices to show $\sup_{\mathcal{F}_n} |G^{SP,K}_n(b) - E[G^{SP,K}_n(b)]| = o_p(1)$ and $\sup_{b\in\mathcal{B}} |G^{SP}(b) - E[G^{SP,K}_n(b)]| = o(1)$, where $\mathcal{F}_n$ denotes the class of functions as $\mathcal{F}_n = \{K(z_1/h_n)\rho(b) : b \in \mathcal{B}\}$.

First, note that $\mathcal{F}_n \subset \mathcal{F} = \{K(z_1/h)\rho(b) : h > 0, b \in \mathcal{B}\} = \mathcal{F}_h \times \mathcal{F}_b$ where $\mathcal{F}_h = \{K(z_1/h) : h > 0\}$ and $\mathcal{F}_b = \{\rho(b) : b \in \mathcal{B}\}$. By Assumption SP8 and Lemma 22(ii) in Nolan and Pollard (1987), $\mathcal{F}_h$ is Euclidean for the constant envelope $\sup_{v\in\mathbb{R}^p} |K(v)| < \infty$. Then, as $\mathcal{F}_b$ is Euclidean for the constant envelope 1 (see Example 2.11 in Pakes and Pollard (1989)), $\mathcal{F}$ is Euclidean for the constant envelope $\sup_{v\in\mathbb{R}^p} |K(v)| < \infty$. Next, note that by Assumptions SP6 and SP8,

$$\begin{aligned}
\sup_{\mathcal{F}_n} E|K(z_1/h_n)\rho(b)| &= \sup_{\mathcal{F}_n} \int E[|K(z_1/h_n)\rho(b)||z_1]f_{z_1}(z_1)dz_1 \\
&= \sup_{\mathcal{F}_n} h^p_n \int |K(v)|E[|\rho(b)||z_1 = vh_n]f_{z_1}(vh_n)dv \\
&\leq \sup_{\mathcal{F}_n} h^p_n \int |K(v)|f_{z_1}(vh_n)dv = O(h^p_n).
\end{aligned}$$

Then, under Assumption SP9, applying Lemma 5 in Honoré and Kyriazidou (2000) yields

$$\sup_{\mathcal{F}_n} h_n^p |G_n^{SP,K}(b) - E[G_n^{SP,K}(b)]| = O_p\left(\sqrt{\frac{h_n^p \log n}{n}}\right) = o_p(h_n^p).$$

As a final step, we show that $\sup_{b\in\mathcal{B}} |G^{SP}(b) - E[G_n^{SP,K}(b)]| = o(1)$. Let $\varphi(\cdot) \equiv f_{z_1}(\cdot)E[\rho(b)|z_1 = \cdot]$ and $\nabla_1\varphi(z)$ denote the gradient of $\varphi(\cdot)$ evaluated at z. Note that by Assumptions SP7–SP9,

$$\begin{aligned}
&\sup_{b\in\mathcal{B}} |G^{SP}(b) - E[G_n^{SP,K}(b)]| \\
=&\sup_{b\in\mathcal{B}} |\varphi(0) - h_n^{-p}\int K(z_1/h_n)\varphi(z_1)dz_1| = \sup_{b\in\mathcal{B}} |\varphi(0) - h_n^{-p}\int K(z_1/h_n)[\varphi(0) + \nabla_1\varphi(\zeta)'z_1]dz_1| \\
=&\sup_{b\in\mathcal{B}} |\varphi(0) - \int K(v)[\varphi(0) + \nabla_1\varphi(v_n)'vh_n]dv| = \sup_{b\in\mathcal{B}} |h_n \int K(v)\nabla_1\varphi(v_n)'vdv| \\
\le& h_n \sup_{b\in\mathcal{B}} \int |K(v)||\nabla_1\varphi(v_n)|_1|v|_1 dv = O(h_n) = o(1),
\end{aligned}$$

where the second equality applies a mean-value expansion and the third equality uses a change of variables. Therefore,

$$\sup_{b\in\mathcal{B}} |G_n^{SP,K}(b) - G^{SP}(b)| \le \sup_{\mathcal{F}_n} |G_n^{SP,K}(b) - E[G_n^{SP,K}(b)]| + \sup_{b\in\mathcal{B}} |G^{SP}(b) - E[G_n^{SP,K}(b)]| = o_p(1),$$

which completes the proof. □

Proof of Theorem 3.2. The proof of the rate of convergence proceeds by verifying the sufficient conditions (Assumptions M (i)–(iii) and D) for applying Lemma 1 of Seo and Otsu (2018). We work with the sample objective function

$$G_n^{SP,K}(b) = \frac{1}{nh_n^{(J-1)p}} \sum_i K_{h_n}(z_{i1})z_{i2} \cdot [sgn(z_{i3}'b) - sgn(z_{i3}'\beta_0)]$$

for the general model with $J+1$ alternatives. Subtracting the term $sgn(z_{i3}'\beta_0)$ is only for the ease of exposition and does not affect the value of the estimator.

First, note that $sgn(z_{i3}'b) - sgn(z_{i3}'\beta_0) = 2(\mathbf{1}[z_{i3}'b \ge 0] - \mathbf{1}[z_{i3}'\beta_0 \ge 0])$. Denote $z_i = (z_{i1}', z_{i2}, z_{i3}')$', $e_n(z_i) = 2h_n^{-(J-1)p}K_{h_n}(z_{i1})z_{i2}$, and $g_{in}(b) = e_n(z_i)(\mathbf{1}[z_{i3}'b \ge 0] - \mathbf{1}[z_{i3}'\beta_0 \ge 0])$. Then, $G_n^{SP,K}(b)$ can be expressed as $n^{-1}\sum_i g_{in}(b)$. Let $\mathcal{N}_0$ denote a neighborhood $\{\beta \in \mathcal{B} : \|\beta - \beta_0\| \le C_0\}$ of β_0 for some constant $C_0 > 0$. As the consistency of $\hat{\beta}$ has been established in Theorem 3.1, $\hat{\beta} \in \mathcal{N}_0$ holds true with probability approaching 1. In what follows, we will suppress the subscript i to simplify the notation as long as this does not cause confusion.

By definition and change of variables, we obtain

$$
\begin{aligned}
E[e_n(z)^2|z_3] &= 4h_n^{-2(J-1)p}E[K(z_1/h_n)^2|z_2 \neq 0, z_3]P(z_2 \neq 0|z_3) \\
&= 4h_n^{-2(J-1)p}\int K(\zeta/h_n)^2 f_{z_1|z_2\neq 0,z_3}(\zeta)d\zeta \cdot P(z_2 \neq 0|z_3) \\
&= 4h_n^{-(J-1)p}\int K(v)^2 f_{z_1|z_2\neq 0,z_3}(vh_n)dv \cdot P(z_2 \neq 0|z_3) \qquad \text{(B.1)}
\end{aligned}
$$

almost surely for all n, and thus by Assumptions SP3, SP6', and SP8', $c_1 < h_n^{(J-1)p}E[e_n(z)^2|z_3] < c_2$ almost surely for some constants $c_1, c_2 > 0$.

Let $\|\cdot\|_2$ denote the $L_2(P)$-norm. We have for all $b_1, b_2 \in \mathcal{N}_0$,

$$
\begin{aligned}
h_n^{(J-1)p/2}\|g_n(b_1) - g_n(b_2)\|_2 &= E[h_n^{(J-1)p}e_n(z)^2(\mathbf{1}[z_3'b_1 \geq 0] - \mathbf{1}[z_3'b_2 \geq 0])^2]^{1/2} \\
&= E[h_n^{(J-1)p}E[e_n(z)^2|z_3](\mathbf{1}[z_3'b_1 \geq 0] - \mathbf{1}[z_3'b_2 \geq 0])^2]^{1/2} \\
&\geq c_1^{1/2}E[|\mathbf{1}[z_3'b_1 \geq 0] - \mathbf{1}[z_3'b_2 \geq 0]|] \geq C_1\|b_1 - b_2\|
\end{aligned}
$$

for some constant $C_1 > 0$, where the last inequality uses the fact that $E[|\mathbf{1}[z_3'b_1 > 0] - \mathbf{1}[z_3'b_2 > 0]|]$ is proportion to the probability for a pair of multi-dimensional wedge shaped regions ($\{z_3'b_1 \geq 0 \geq z_3'b_2\} \cup \{z_3'b_2 \geq 0 \geq z_3'b_1\}$) with an angle of order $\|b_1 - b_2\|$ (see the discussion in Kim and Pollard (1990) p. 214). This verifies Assumption M(ii) in Seo and Otsu (2018).

We use similar arguments to obtain that for some $\beta \in \mathcal{N}_0$ and small $\varepsilon > 0$

$$
\begin{aligned}
& h_n^{(J-1)p}E[\sup_{b\in\mathcal{B}:\|b-\beta\|<\varepsilon}|g_n(b) - g_n(\beta)|^2] \\
=& E[h_n^{(J-1)p}E[e_n(z)^2|z_3]\sup_{b\in\mathcal{B}:\|b-\beta\|<\varepsilon}|\mathbf{1}[z_3'b \geq 0] - \mathbf{1}[z_3'\beta \geq 0)|] \\
\leq& c_2E[\sup_{b\in\mathcal{B}:\|b-\beta\|<\varepsilon}|\mathbf{1}[z_3'b \geq 0] - \mathbf{1}[z_3'\beta \geq 0]|] \leq C_2\varepsilon \qquad \text{(B.2)}
\end{aligned}
$$

for some $C_2 > 0$ and n large enough. This verifies Assumption M(iii) in Seo and Otsu (2018).

Next, note that under Assumptions SP7'–SP9', a change of variables yields

$$
\begin{aligned}
E[g_n(b)] =& h_n^{-(J-1)p}\int K(\zeta/h_n)E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = \zeta]f_{z_1}(\zeta)d\zeta \\
=& \int K(v)E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = vh_n]f_{z_1}(vh_n)dv \\
=& f_{z_1}(0)E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = 0] \\
&+ h_n^2\int K(v)v'\frac{\partial^2 f_{z_1}(\tau)E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = \tau]}{\partial\tau\partial\tau'}|_{\tau=\bar{v}}vdv, \qquad \text{(B.3)}
\end{aligned}
$$

where $\bar{v}$ is a point on the line joining 0 and vh_n, and the third equality follows from $\int vK(v)dv = 0$, the dominated convergence theorem, and the mean value theorem. Assumptions SP7' and SP9' imply that the last term in (B.3) is $o((nh_n^{(J-1)p})^{-2/3})$. Then, we write

$$\begin{aligned}E[g_n(b)] =& f_{z_1}(0)E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = 0] + o((nh_n^{(J-1)p})^{-2/3})\\ =& f_{z_1}(0)\left((b-\beta_0)'\frac{\partial}{\partial b}E[z_2(sgn(z_3'b)|z_1=0]|_{b=\beta_0}\right.\\ &\left.+\frac{1}{2}(b-\beta_0)'\frac{\partial^2 E[z_2(sgn(z_3'b))|z_1=0]}{\partial b\partial b'}|_{b=\beta_0}(b-\beta_0)\right)\\ &+ o_p(\|b-\beta_0\|^2) + o((nh_n^{(J-1)p})^{-2/3}. \end{aligned} \tag{B.4}$$

As deduced in Lemma B.1, we have

$$\begin{aligned}-E[z_2(sgn(z_3'b) - sgn(z_3'\beta_0))|z_1 = 0] &= 2\int_{\mathcal{Z}_b} sgn(z_3'\beta_0)E[z_2|z_3, z_1 = 0]dF_{z_3|z_1=0}\\ &= 2\int_{\mathcal{Z}_b} |E[z_2|z_3, z_1 = 0]|dF_{z_3|z_1=0} > 0\end{aligned}$$

holds true for all $b \in \mathcal{B} \setminus \{\beta_0\}$. Then, applying the same argument as Kim and Pollard (1990) pp. 214 - 215 yields

$$\frac{\partial}{\partial b}E[z_2(sgn(z_3'b))|z_1 = 0]|_{b=\beta_0} = 0 \tag{B.5}$$

and

$$\begin{aligned}&\frac{\partial^2 E[z_2(sgn(z_3'b))|z_1 = 0]}{\partial b\partial b'}|_{b=\beta_0}\\ =& -\int \mathbf{1}[z_3'\beta_0 = 0]\left(\frac{\partial}{\partial z_3}E[z_2|z_3, z_1 = 0]\right)'\beta_0 z_3 z_3' f_{z_3|z_1=0}(z_3)d\mu_{\beta_0},\end{aligned} \tag{B.6}$$

where μ_{β_0} is the surface measure on the boundary of $\{z_3 : z_3'\beta_0 \geq 0\}$.

Combine (B.4), (B.5), and (B.6) to write

$$E[g_n(b)] = \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + o_p(\|b-\beta_0\|^2) + o((nh_n^{(J-1)p})^{-2/3}), \tag{B.7}$$

where the matrix

$$V = -f_{z_1}(0)\int \mathbf{1}[z_3'\beta_0 = 0]\left(\frac{\partial}{\partial z_3}E[z_2|z_3, z_1 = 0]\right)'\beta_0 z_3 z_3' f_{z_3|z_1=0}(z_3)d\mu_{\beta_0} \tag{B.8}$$

is negative definite. This verifies Assumption M(i) in Seo and Otsu (2018).

Notice that $h_n^{(J-1)p}g_n(b)$ is uniformly bounded by Assumption SP8' and $\lim_{n\to\infty} E[g_n(b)]$ is uniquely maximized at β_0 by Lemma B.1. Besides, Assumption D in Seo and Otsu (2018) is satisfied trivially under Assumption SP1. Then, by Lemma 1 of Seo and Otsu (2018), we conclude that there exists some positive constant C for each $\varepsilon > 0$ such that

$$\left| \frac{1}{n} \sum_i g_{in}(b) - E[g_n(b)] \right| \leq \varepsilon \|b - \beta_0\|^2 + O_p\left(\left(nh_n^{(J-1)p}\right)^{-2/3}\right) \tag{B.9}$$

for all $b \in \{\beta \in \mathcal{B} : (nh_n^{(J-1)p})^{-1/3} \leq \|\beta - \beta_0\| \leq C\}$. Then, assuming $\|\hat{\beta} - \beta_0\| \geq (nh_n^{(J-1)p})^{-1/3}$, we have, by (B.7) and (B.9),

$$\begin{aligned} \frac{1}{n} \sum_i g_{in}(\hat{\beta}) &\leq E[g_n(\hat{\beta})] + \varepsilon \|\hat{\beta} - \beta_0\|^2 + O_p((nh_n^{(J-1)p})^{-2/3}) \\ &\leq (\varepsilon - C_V)\|\hat{\beta} - \beta_0\|^2 + o(\|\hat{\beta} - \beta_0\|^2) + O_p((nh_n^{(J-1)p})^{-2/3}) \end{aligned} \tag{B.10}$$

for each $\varepsilon > 0$ and a positive constant C_V (determined by V). By the definitions of $\hat{\beta}$ and $g_{in}(\cdot)$,

$$\begin{aligned} \frac{1}{n} \sum_i g_{in}(\hat{\beta}) &\geq \sup_{b \in \mathcal{B}} \frac{1}{n} \sum_i g_{in}(b) - o_p((nh_n^{(J-1)p})^{-2/3}) \geq \frac{1}{n} \sum_i g_{in}(\beta_0) - o_p((nh_n^{(J-1)p})^{-2/3}) \\ &= o_p((nh_n^{(J-1)p})^{-2/3}). \end{aligned} \tag{B.11}$$

Combine (B.10) and (B.11) to deduce

$$o_p((nh_n^{(J-1)p})^{-2/3}) \leq (\varepsilon - C_V)\|\hat{\beta} - \beta_0\|^2 + o(\|\hat{\beta} - \beta_0\|^2) + O_p((nh_n^{(J-1)p})^{-2/3}).$$

Then, $\hat{\beta} - \beta_0 = O_p((nh_n^{(J-1)p})^{-1/3})$ follows from taking ε sufficiently small such that $\varepsilon - C_V < 0$.

Given the rate result, the final step is to establish the limiting distribution of $\hat{\beta}$. We do this by applying Theorem 1 of Seo and Otsu (2018), which extends Theorem 2.7 of Kim and Pollard (1990) (a continuous mapping theorem of an argmax element) to the case where the objective function can vary with the sample size. To this end, it suffices to verify the sufficient conditions of Theorem 2.3 of Kim and Pollard (1990) to establish the weak convergence of the following (normalized empirical process)

$$Z_n(\mathbf{s}) = n^{1/6} h_n^{2(J-1)p/3} \mathbb{G}_n(g_n(\beta_0 + \mathbf{s}(nh_n^{(J-1)p})^{-1/3}) - g_n(\beta_0))$$

for $\mathbf{s} \in \mathbb{R}^p$ with $\|\mathbf{s}\| < \infty$, where $\mathbb{G}_n g(b) \equiv \sqrt{n}(n^{-1} \sum_i g_{in}(b) - E[g_n(b)])$ for all $b \in \mathcal{B}$. This involves checking the the finite-dimensional convergence and stochastic asymptotic equicontinuity of Z_n. Denote $\Psi_{n,\mathbf{s}}(z_3) = \mathbf{1}[z_3'(\beta_0 + \mathbf{s}(nh_n^{(J-1)p})^{-1/3}) \geq 0] - \mathbf{1}[z_3'\beta_0 \geq 0]$ and $g_{n,\mathbf{s}} = n^{1/6} h_n^{2(J-1)p/3}[g_n(\beta_0 + \mathbf{s}(nh_n^{(J-1)p})^{-1/3}) - g_n(\beta_0)] = n^{1/6} h_n^{2(J-1)p/3} e_n(z) \Psi_{n,\mathbf{s}}(z_3)$.

We use the central limit theorem in Lemma C of Seo and Otsu (2018) to establish the finite-dimensional convergence of Z_n. Given all the results at hand, this reduces to verifying a sufficient Lindeberg-type condition therein. Note that for some positive constant M and bounded set $\mathcal{M}$ in

$\mathbb{R}^{(J-1)p}$,

$$\begin{aligned}P(|g_{n,\mathbf{s}}| \geq M) &= P(2n^{1/6}h_n^{-(J-1)p/3}|K_{h_n}(z_1)z_2| \geq M||\Psi_{n,\mathbf{s}}(z_3)| = 1)P(|\Psi_{n,\mathbf{s}}(z_3)| = 1)\\ &\leq P(|K_{h_n}(z_1)| \geq Mn^{-1/6}h_n^{(J-1)p/3}/2||\Psi_{n,\mathbf{s}}(z_3)| = 1)P(|\Psi_{n,\mathbf{s}}(z_3)| = 1)\\ &\leq P(z_1/h_n \in \mathcal{M}||\Psi_{n,\mathbf{s}}(z_3)| = 1)P(|\Psi_{n,\mathbf{s}}(z_3)| = 1) = O((nh_n^{-2(J-1)p})^{-1/3}),\end{aligned}$$

where the first equality uses the fact that $|\Psi_{n,\mathbf{s}}(z_3)|$ takes only 0 or 1, the first inequality is due to $|z_2| \leq 1$, the second inequality follows from the bounded support of $K(\cdot)$, and the last equality is the result of $P(z_1/h_n \in \mathcal{M}||\Psi_{n,\mathbf{s}}(z_3)| = 1) = O(h_n^{(J-1)p})$ (by the boundedness of the density of z_1) and $P(|\Psi_{n,\mathbf{s}}(z_3)| = 1) = O((nh_n^{(J-1)p})^{-1/3})$. By Lemma 2 of Seo and Otsu (2018), this result is sufficient for the Lindeberg-type condition and hence for applying Lemma C of Seo and Otsu (2018) to conclude the finite-dimensional convergence of Z_n.

Finally, by definition,

$$\begin{aligned}\|g_{n,\mathbf{s}}\|_2 =& n^{1/6}h_n^{2(J-1)p/3}\sqrt{E[E[e_n(z)^2|z_3]\Psi_{n,\mathbf{s}}(z_3)^2]} \leq n^{1/6}h_n^{(J-1)p/6}E[|\Psi_{n,\mathbf{s}}(z_3)|]\\ =& O((nh_n^{(J-1)p})^{-1/6}),\end{aligned} \tag{B.12}$$

where the first inequality follows from (B.1) and the last equality uses the fact that $E[|\Psi_{n,\mathbf{s}}(z_3)|] = O((nh_n^{(J-1)p})^{-1/3})$. Then, (B.2) and (B.12) together are sufficient for invoking Lemma M' of Seo and Otsu (2018) to establish the stochastic asymptotic equicontinuity of Z_n.

Collecting all these results, we conclude, by Theorem 1 of Seo and Otsu (2018), that the limiting distribution of $\hat{\beta}$ is of the form as in Theorem 3.2. The matrix V is given in (B.8). The covariance kernel H can be obtained in the same way as in Kim and Pollard (1990) (p. 215). That is, decompose z_3 into $\xi'\beta_0 + z_3^\perp$ with $z_3^\perp$ orthogonal to β_0. Then we write

$$H(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2}\left(L(\mathbf{s}_1) + L(\mathbf{s}_2) - L(\mathbf{s}_1 - \mathbf{s}_2)\right), \tag{B.13}$$

where $L(\mathbf{s}) = f_{z_1}(0)\int |z_3^{\perp\prime}\mathbf{s}|p(0, z_3^\perp|z_1 = 0)dz_3^\perp$ with $p(\cdot, \cdot|z_1 = 0)$ being joint density of $(\xi, z_3^\perp)$ conditional on $z_1 = 0$. □

B.2 Dynamic Panel Data Estimator

Lemmas B.2–B.4 establish the identification of θ_0 in the dynamic model, based on which we show the consistency of $\hat{\theta}$ in Lemma B.5. The proofs of the rate of convergence and asymptotic distribution of $\hat{\theta}$ are omitted as the derivation invokes essentially the same arguments used for proving Theorem 3.2.

Throughout this section, we work with sample objective function

$$G_n^{DP,K}(\theta) \equiv \frac{1}{nh_n^{3p}} \sum_i K_{h_n}(z_{1i}) z_{2i} \cdot sgn(z_{3i}'\theta)$$

and population objective function $G^{DP}(\theta) \equiv f_{z_1}(0)E[\psi(\theta)|z_1 = 0]$. As in Section B.1, we assume $\beta_0^{(1)} > 0$ w.l.o.g. as the case $\beta_0^{(1)} < 0$ is symmetric. Besides, we define events $\Omega = \{z_1 = 0\}$, $A = \{y_{10} = d_0, y_{11} = 1, y_{12} = 0, y_{13} = d_3\}$, and $B = \{y_{10} = d_0, y_{11} = 0, y_{12} = 1, y_{13} = d_3\}$, where $(d_0, d_3) \in \{0,1\}^2$.

Lemma B.2. *If Assumptions DP1–DP3 holds,* $sgn(P(A|x,\alpha,\Omega) - P(B|x,\alpha,\Omega)) = sgn(z_3'\theta_0)$.

Proof of Lemma B.2. By Assmption DP3, we write

$$\begin{aligned} P(A|x,\alpha,\Omega) =& P(y_{10} = 1|x,\alpha,\Omega)^{d_0}(1 - P(y_{10} = 1|x,\alpha,\Omega))^{1-d_0} \\ &\times P(x_{11}'\beta_0 + \gamma_0 d_0 + \alpha_1 - \epsilon_{11} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{21}, 0\}|x,\alpha,\Omega) \\ &\times (1 - P(x_{12}'\beta_0 + \gamma_0 + \alpha_1 - \epsilon_{12} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{22}, 0\}|x,\alpha,\Omega)) \\ &\times P(x_{12}'\beta_0 + \alpha_1 - \epsilon_{13} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{23}, 0\}|x,\alpha,\Omega)^{d_3} \\ &\times (1 - P(x_{12}'\beta_0 + \alpha_1 - \epsilon_{13} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{23}, 0\}|x,\alpha,\Omega))^{1-d_3}, \end{aligned}$$

and similarly,

$$\begin{aligned} P(B|x,\alpha,\Omega) =& P(y_{10} = 1|x,\alpha,\Omega)^{d_0}(1 - P(y_{10} = 1|x,\alpha,\Omega))^{1-d_0} \\ &\times (1 - P(x_{11}'\beta_0 + \gamma_0 d_0 + \alpha_1 - \epsilon_{11} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{21}, 0\}|x,\alpha,\Omega)) \\ &\times P(x_{12}'\beta_0 + \alpha_1 - \epsilon_{12} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{22}, 0\}|x,\alpha,\Omega) \\ &\times P(x_{12}'\beta_0 + \gamma_0 + \alpha_1 - \epsilon_{13} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{23}, 0\}|x,\alpha,\Omega)^{d_3} \\ &\times (1 - P(x_{12}'\beta_0 + \gamma_0 + \alpha_1 - \epsilon_{13} > \max\{x_{21}'\beta_0 + \alpha_2 - \epsilon_{23}, 0\}|x,\alpha,\Omega))^{1-d_3}. \end{aligned}$$

It is not hard to verify that

$$\frac{P(A|x,\alpha,\Omega)}{P(B|x,\alpha,\Omega)} \geq 1 \Leftrightarrow x_{11}'\beta_0 + \gamma_0 d_0 \geq x_{12}'\beta_0 + \gamma_0 d_3.$$

for each of the 4 cases corresponding to the values of d_0 and d_3. Then the desired result follows. □

Lemma B.3. *If Assumptions DP1–DP5 hold,* $P(sgn(z_3'\theta) \neq sgn(z_3'\theta_0)|\Omega) > 0$ *for all* $\theta \in \Theta \setminus \{\theta_0\}$.

Proof of Lemma B.3. The statement in the lemma is equivalent to

$$P(x_{1(12)}b^{(1)} + \tilde{z}_3'\tilde{\theta} > 0 > x_{1(12)}\beta_0^{(1)} + \tilde{z}_3'\tilde{\theta}_0|\Omega) + P(x_{1(12)}\beta_0^{(1)} + \tilde{z}_3'\tilde{\theta}_0 > 0 > x_{1(12)}b^{(1)} + \tilde{z}_3'\tilde{\theta}|\Omega) > 0.$$

Note that by Assumption DP4, this statement holds true whenever $b^{(1)} < 0$, and hence we focus on the case $b^{(1)} > 0$ in what follows.

To prove the statement above, it suffices to show that for all $\theta \in \Theta \setminus \{\theta_0\}$, (i) $P(\tilde{z}_3'\tilde{\theta}/b^{(1)} \neq \tilde{z}_3'\tilde{\theta}_0/\beta_0^{(1)}|\Omega) > 0$, and (ii) $P(x_{1(12)}^{(1)} \in \mathcal{I}|\tilde{x}_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega) > 0$ for all $(d_0, d_3) \in \{0,1\}^2$ and for any proper interval $\mathcal{I}$ on the real line.

We start from proving statement (i). Note that if $r/b^{(1)} = \gamma_0/\beta_0^{(1)}$, $P(\tilde{z}_3'\tilde{\theta}/b^{(1)} = \tilde{z}_3'\tilde{\theta}_0/\beta_0^{(1)}|\Omega) = P(\tilde{x}_{1(12)}'(\tilde{b}/b^{(1)} - \tilde{\beta}_0/\beta_0^{(1)}) = 0|\Omega)$. In this case, $\tilde{b}/b^{(1)} \neq \tilde{\beta}_0/\beta_0^{(1)}$ must hold, for otherwise $\theta = \theta_0$ holds true. To see this, note that $\tilde{b}/b^{(1)} = \tilde{\beta}_0/\beta_0^{(1)}$, together with $r/b^{(1)} = \gamma_0/\beta_0^{(1)}$, implies $\beta_0^{(1)}\theta = b^{(1)}\theta_0$, which further implies $b^{(1)} = \beta_0^{(1)}$ as $\|\theta\| = \|\theta_0\|$ is assumed. Then statement (i) follows by Assumption DP5.

For the case with $r/b^{(1)} \neq \gamma_0/\beta_0^{(1)}$, we write

$$
\begin{aligned}
& P(\tilde{z}_3'\tilde{\theta}/b^{(1)} = \tilde{z}_3'\tilde{\theta}_0/\beta_0^{(1)}|\Omega) \\
= & \sum_{d_0 \in \{0,1\}} \int P(y_{13} = d_0 + \tilde{x}_{1(12)}'(\tilde{b}/b^{(1)} - \tilde{\beta}_0/\beta_0^{(1)})/(r/b^{(1)} - \gamma_0/\beta_0^{(1)})|y_{10} = d_0, \tilde{x}_{1(12)}, \Omega) \\
& \times P(y_{10} = d_0|\tilde{x}_{1(12)}, \Omega) dF_{\tilde{x}_{1(12)}|\Omega},
\end{aligned}
$$

where $P(y_{13} = d_0 + \tilde{x}_{1(12)}'(\tilde{b}/b^{(1)} - \tilde{\beta}_0/\beta_0^{(1)})/(r/b^{(1)} - \gamma_0/\beta_0^{(1)})|y_{10} = d_0, \tilde{x}_{1(12)}, \Omega) < 1$ holds true for all $d_0 \in \{0,1\}$ by Assumption DP3. This implies $P(\tilde{z}_3'\tilde{\theta}/b^{(1)} \neq \tilde{z}_3'\tilde{\theta}_0/\beta_0^{(1)}|\Omega) > 0$.

We now move on to the proof of statement (ii). We write, by Bayes rule,

$$
\begin{aligned}
& P(x_{1(12)}^{(1)} \in \mathcal{I}|\tilde{x}_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega) \\
= & \frac{P(y_{10} = d_0, y_{13} = d_3|\tilde{x}_{1(12)}, x_{1(12)}^{(1)} \in \mathcal{I}, \Omega) P(x_{1(12)}^{(1)} \in \mathcal{N}|\tilde{x}_{1(12)}, \Omega)}{P(y_{10} = d_0, y_{13} = d_3|\tilde{x}_{1(12)}, \Omega)}.
\end{aligned}
$$

Assumption DP4 secures that $P(x_{1(12)}^{(1)} \in \mathcal{I}|\tilde{x}_{1(12)}, \Omega) > 0$. Furthermore, note that

$$
\begin{aligned}
& P(y_{10} = d_0, y_{13} = d_3|\tilde{x}_{1(12)}, x_{1(12)}^{(1)} \in \mathcal{I}, \Omega) \\
= & \int P(y_{13} = d_3|x, \alpha, y_{10} = d_0, \Omega) P(y_{10} = d_0|x, \alpha, \Omega) dF_{x,\alpha|\tilde{x}_{1(12)}, x_{1(12)}^{(1)} \in \mathcal{I}, \Omega} \\
= & \sum_{(d_1,d_2) \in \{0,1\}^2} \int P(y_{13} = d_3|x, \alpha, y_{10} = d_0, y_{11} = d_1, y_{12} = d_2, \Omega) \\
& \times P(y_{12} = d_2|x, \alpha, y_{10} = d_0, y_{11} = d_1, \Omega) P(y_{11} = d_1|x, \alpha, y_{10} = d_0, \Omega) \\
& \times P(y_{10} = d_0|x, \alpha, \Omega) dF_{x,\alpha|\tilde{x}_{1(12)}, x_{1(12)}^{(1)} \in \mathcal{I}, \Omega}.
\end{aligned}
$$

Therefore, $P(y_{10} = d_0, y_{13} = d_3|\tilde{x}_{1(12)}, x_{1(12)}^{(1)} \in \mathcal{I}, \Omega) > 0$ by Assumption DP3, and thus $P(x_{1(12)}^{(1)} \in$

$\mathcal{I}|\tilde{x}_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega) > 0$, which completes the proof. □

Lemma B.4. *If Assumptions DP1–DP6 hold, $G^{DP}(\theta_0) > G^{DP}(\theta)$ for all $\theta \in \Theta \setminus \{\theta_0\}$.*

Proof of Lemma B.4. Recall that $\psi(\theta) = z_2 \cdot sgn(z_3'\theta)$ for all $\theta \in \Theta$. Let $\mathcal{Z}_\theta \equiv \{z_3 : sgn(z_3'\theta) \neq sgn(z_3'\theta_0)\}$. Lemma B.3 shows that $P(\mathcal{Z}_\theta|\Omega) > 0$ holds true for all $\theta \in \Theta \setminus \{\theta_0\}$. Then, by definition,

$$
\begin{aligned}
& G^{DP}(\theta_0) - G^{DP}(\theta) \\
=& f_{z_1}(0) E[z_2 \cdot (sgn(z_3'\theta_0) - sgn(z_3'\theta))|\Omega] = 2 f_{z_1}(0) \int_{\mathcal{Z}_\theta} sgn(z_3'\theta_0) E[z_2|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_\theta} sgn(z_3'\theta_0) E[E[y_{1(12)}|x, \alpha, y_{10} = d_0, y_{13} = d_3, \Omega]|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_\theta} sgn(z_3'\theta_0) E[E[\mathbf{1}[y_{11} = 1, y_{12} = 0]|x, \alpha, y_{10} = d_0, y_{13} = d_3, \Omega] \\
& - E[\mathbf{1}[y_{11} = 0, y_{12} = 1]|x, \alpha, y_{10} = d_0, y_{13} = d_3, \Omega]|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_\theta} sgn(z_3'\theta_0) E[P(y_{11} = 1, y_{12} = 0|x, \alpha, y_{10} = d_0, y_{13} = d_3, \Omega) \\
& - P(y_{11} = 0, y_{12} = 1|x, \alpha, y_{10} = d_0, y_{13} = d_3, \Omega)|z_3, \Omega] dF_{z_3|\Omega} \\
=& 2 f_{z_1}(0) \int_{\mathcal{Z}_\theta} E\left[sgn(z_3'\theta_0) \left(\frac{P(A|x, \alpha, \Omega) - P(B|x, \alpha, \Omega)}{P(y_{10} = d_0, y_{13} = d_3|x, \alpha, \Omega)} \right) |z_3, \Omega \right] dF_{z_3|\Omega}.
\end{aligned}
$$

It follows from Lemma B.2 that

$$
sgn(z_3'\theta_0) \left(\frac{P(A|x, \alpha, \Omega) - P(B|x, \alpha, \Omega)}{P(y_{10} = d_0, y_{13} = d_3|x, \alpha, \Omega)} \right) \geq 0,
$$

and hence

$$
E\left[sgn(z_3'\theta_0) \left(\frac{P(A|x, \alpha, \Omega) - P(B|x, \alpha, \Omega)}{P(y_{10} = d_0, y_{13} = d_3|x, \alpha, \Omega)} \right) |z_3, \Omega \right] = E\left[\left| \frac{P(A|x, \alpha, \Omega) - P(B|x, \alpha, \Omega)}{P(y_{10} = d_0, y_{13} = d_3|x, \alpha, \Omega)} \right| |z_3, \Omega \right].
$$

The expectation above is strictly positive with probability 1 as $P(A|x, \alpha, \Omega) - P(B|x, \alpha, \Omega) = 0$ if and only if $sgn(z_3'\theta_0) = 0$ which is an event having zero probability measure under Assumption DP4. Then the desired result follows from Lemma B.3 and Assumption DP6. □

Lemma B.5. *If Assumptions DP1–DP9 hold, $\hat{\theta} \xrightarrow{p} \theta_0$.*

Proof of Lemma B.5. The proof is standard, which proceeds by verifying the four sufficient conditions for applying Theorem 2.1 in Newey and McFadden (1994): (C1) Θ is a compact set, (C2) $\sup_{\theta \in \Theta} |G_n^{DP,K}(\theta) - G^{DP}(\theta)| = o_p(1)$, (C3) $G^{DP}(\theta)$ is continuous in θ, and (C4) $G^{DP}(\theta)$ is uniquely maximized at θ_0.

The compactness of Θ is satisfied by Assumption DP2. The identification condition (C4) is established by Lemma B.4. The verification of the uniform convergence condition (C2) is omitted as it follows from identical arguments to those used for proving Theorem 3.1. The remaining task is to check the continuity condition (C3). Note that we can express $G^{DP}(\theta)$ as the sum of functions (with respect to θ) of the following form:

$$
\begin{aligned}
& P(y_{11} - y_{12} = d, x^{(1)}_{1(12)} b^{(1)} + \tilde{x}'_{1(12)} \tilde{b} + g(y_{10} - y_{13}) > 0 | \Omega) \\
= & \sum_{(d_0, d_3) \in \{0,1\}^2} \int \left[\int_{-(\tilde{x}'_{1(12)} \tilde{b} + r(d_0 - d_3))/b^{(1)}}^{\infty} P(y_{11} - y_{12} = d | x_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega) \right. \\
& \left. \times f_{x^{(1)}_{1(12)} | \tilde{x}_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega}(x) dx \right] dF_{\tilde{x}_{1(12)} | y_{10} = d_0, y_{13} = d_3, \Omega} \times P(y_{10} = d_0, y_{13} = d_3 | \Omega)
\end{aligned}
$$

for some $d \in \{-1, 0, 1\}$. Note that to secure continuity of the function of this form with respect to θ it is sufficient that $f_{x^{(1)}_{1(12)} | \tilde{x}_{1(12)}, y_{10} = d_0, y_{13} = d_3, \Omega}(\cdot)$ does not have any mass points, which is implied by Assumptions DP3–DP4 and Bayes rule. □

C Additional Simulation Results

This appendix contains the results of additional simulations. The five designs (Designs 1C–5C) we examine here are counterparts of Designs 1–5 in Section 4, respectively, but using all continuous regressors. Specifically, we consider

- Design 1C: $(x^{(1)}_{ij})_{i=1,\ldots,n; j=1,2}$ are i.i.d. $N(0,1)$ random variables, and all other $x^{(l)}_{ij}$ for $i = 1, \ldots, n; j = 1, 2; l = 2, 3$ are i.i.d.uniform random variables in $[-1, 1]$.
- Design 2C: $(x^{(1)}_{ij})_{i=1,\ldots,n; j=1,2}$ are i.i.d. $N(0,1)$ random variables, and all other $x^{(l)}_{ij}$ for $i = 1, \ldots, n; j = 1, 2; l = 2, 3, 4, 5$ are i.i.d. uniform random variables in $[-1, 1]$.
- Design 3C: $(x^{(1)}_{ij})_{i=1,\ldots,n; j=1,2,3,4}$ are i.i.d. $N(0,1)$ random variables, and all other $x^{(l)}_{ij}$ for $i = 1, \ldots, n; j = 1, 2, 3, 4; l = 2, 3$ are i.i.d. uniform random variables in $[-1, 1]$.
- Design 4C: $(x^{(1)}_{ijt})_{i=1,\ldots,n; j=1,2; t=1,2}$ are i.i.d. $N(0,1)$ random variables, all other $x^{(l)}_{ijt}$ for $i = 1, \ldots, n; j = 1, 2; t = 1, 2; l = 2, 3$ are i.i.d. uniform random variables in $[-1, 1]$, and $\alpha_{ij} = (x^{(1)}_{ij1} + x^{(1)}_{ij2})/4$ for $j = 1, 2$.
- Design 5C: $(x^{(1)}_{ijt})_{i=1,\ldots,n; j=1,2; t=0,1,2,3}$ are i.i.d. $N(0,1)$ random variables, all other $x^{(l)}_{ijt}$ for $i = 1, \ldots, n; j = 1, 2; t = 0, 1, 2, 3; l = 2, 3$ are i.i.d. uniform random variables in $[-1, 1]$, and $\alpha_{ij} = (x^{(1)}_{ij0} + x^{(1)}_{ij1} + x^{(1)}_{ij2} + x^{(1)}_{ij3})/8$ for $j = 1, 2$.

The results for Designs 1C–5C are reported below in Tables 1C–7C corresponding to Tables 1–7 in Section 4, respectively.

Table 1C: (Design 1C) Cross-Sectional Design with $J = 2$ and $p = 3$

	β_1				β_2			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 250$	0.0180	0.2067	0.0062	0.1420	0.0204	0.2128	0.0162	0.1413
$n = 500$	0.0052	0.1427	0.0029	0.0987	0.0017	0.1490	-0.0105	0.0983
$n = 1000$	0.0069	0.0998	0.0040	0.0675	0.0083	0.1058	0.0046	0.0705

Table 2C: (Design 2C) Cross-Sectional Design with $J = 2$ and $p = 5$

	β_1				β_2			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 250$	0.0355	0.2302	0.0199	0.1505	0.0226	0.2324	0.0113	0.1459
$n = 500$	0.0089	0.1589	-0.0023	0.1057	0.0120	0.1559	0.0069	0.1059
$n = 1000$	0.0073	0.1146	0.0057	0.0762	0.0076	0.1120	0.0061	0.0742

Table 3C: (Design 3C) Cross-Sectional Design with $J = 4$ and $p = 3$

	β_1				β_2			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 250$	0.0223	0.2519	0.0094	0.1600	0.0150	0.2577	-0.0069	0.1640
$n = 500$	0.0155	0.1762	0.0102	0.1126	0.0172	0.1757	0.0086	0.1168
$n = 1000$	0.0107	0.1251	0.0062	0.0878	0.0094	0.1340	0.0088	0.0903

Table 4C: (Design 4C) Static Panel Design with $J = 2$, $p = 3$, and $t \in \{1, 2\}$

	β_1				β_2			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 500$	0.0256	0.3180	0.0053	0.2219	0.0227	0.3152	0.0070	0.2092
$n = 1000$	0.0268	0.2655	0.0230	0.1703	0.0257	0.2568	0.0153	0.1624
$n = 2000$	0.0186	0.2166	0.0087	0.1448	0.0250	0.2185	0.0174	0.1485
$n = 5000$	0.0106	0.1734	0.0078	0.1158	0.0108	0.1761	0.0002	0.1168
$n = 10000$	0.0064	0.1421	-0.0016	0.0956	0.0150	0.1468	0.0096	0.0983

Table 5C: (Design 4C, Two-step) Static Panel Design with $J = 2$, $p = 3$, and $t \in \{1, 2\}$

	β_1				β_2			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 500$	0.0126	0.3188	-0.0152	0.2077	0.0189	0.3164	-0.0030	0.2112
$n = 1000$	0.0199	0.2551	0.0012	0.1729	0.0225	0.2615	0.0028	0.1659
$n = 2000$	0.0191	0.2084	0.0052	0.1344	0.0130	0.2145	-0.0012	0.1446
$n = 5000$	0.0018	0.1506	-0.0024	0.0944	0.0067	0.1623	-0.0014	0.1081
$n = 10000$	0.0073	0.1247	0.0010	0.0898	0.0101	0.1266	0.0063	0.0917

Table 6C: (Design 5C) Dynamic Panel Design with $J = 2$, $p = 3$, and $t \in \{0, 1, 2, 3\}$

	β				γ			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 500$	0.0483	0.3911	0.0066	0.2607	-0.0352	0.2966	-0.0384	0.2549
$n = 1000$	0.0560	0.3466	0.0263	0.2279	-0.0120	0.2863	-0.0157	0.2363
$n = 2000$	0.0430	0.2957	0.0130	0.1814	-0.0265	0.2822	-0.0285	0.2313
$n = 5000$	0.0179	0.2509	-0.0128	0.1601	-0.0056	0.2590	-0.0101	0.1999
$n = 10000$	0.0242	0.2191	0.0117	0.1472	-0.0024	0.2460	0.0045	0.1858

Table 7C: (Design 5C, Two-step) Dynamic Panel Design with $J = 2$, $p = 3$, and $t \in \{0, 1, 2, 3\}$

	β				γ			
	MEAN	RMSE	MED	MAE	MEAN	RMSE	MED	MAE
$n = 500$	0.0492	0.3292	0.0100	0.2061	-0.0430	0.2915	-0.0531	0.2385
$n = 1000$	0.0405	0.2804	-0.0020	0.1805	-0.0001	0.2816	0.0103	0.2268
$n = 2000$	0.0254	0.2236	-0.0051	0.1408	-0.0136	0.2716	-0.0048	0.2198
$n = 5000$	0.0105	0.1743	-0.0026	0.1149	-0.0067	0.2456	-0.0065	0.1894
$n = 10000$	0.0107	0.1446	-0.0007	0.0974	-0.0168	0.2290	-0.0130	0.1595

References

Honoré, B. and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

Kim, J. and D. Pollard (1990): "Cube root asymptotics," *The Annals of Statistics*, 18, 191–219.

Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle and D. McFadden, North Holland.

Nolan, D. and D. Pollard (1987): "U-Processes: Rates of Convergence," *Annals of Statistics*, 15, 780–799.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Seo, M. and T. Otsu (2018): "Local M-estimation with discontinuous criterion for dependent and limited observations," *Annals of Statistics*, 46, 344–369.

Serfling, R. J. (2009): *Approximation theorems of mathematical statistics*, vol. 162, John Wiley & Sons.

Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

——— (1994a): "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators," *Annals of Statistics*, 22, 439–459.

——— (1994b): "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator," *Econometric Theory*, 10, 372–395.