Supplement to "SEMIPARAMETRIC ESTIMATION OF DYNAMIC BINARY CHOICE PANEL DATA MODELS"

Fu Ouyang*[1] and Thomas Tao Yang†[2]

[1]School of Economics, University of Queensland

[2]Research School of Economics, Australian National University

February 9, 2024

Appendix D contains proofs for all technical lemmas used in Appendices A and B. Appendix E offers technical details for Section 5. Additionally, Appendix F presents tables summarizing simulation results for Designs 3–5. It also includes supplementary Monte Carlo experiments (Designs 6–8) that investigate the impact of serial correlations of x_{it} on our proposed estimation and inference procedure.

D Proofs for Technical Lemmas

Proof of Lemma A.1. Here we only prove the case $\tau = s$. The derivation for case $\tau = t$ is analogous.

First, note that by the law of total probability, we can write for all $d_1 \in \{0, 1\}$,

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha)\\
=&\sum_{j=1}^{3}\left\{P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha, E_{s+1,j})\right.\\
&\left.\times P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = d_1, \alpha)\right\}. \qquad \text{(D.1)}
\end{aligned}$$

When $d_1 = 1$, (D.1) reduces to

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha)\\
=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,1})\\
&\times P(E_{s+1,1}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \qquad \text{(D.2)}
\end{aligned}$$

*Email: f.ouyang@uq.edu.au.

†Email: tao.yang@anu.edu.au.

as by definition $E_{s+1,3} \cap \{y_{s+1} = 1\} = \emptyset$ and by Bayes' theorem[1]

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,2}) \\
=&\frac{P(y_{s+1} = 1, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha)}{P(y_{s+1} = 1, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha)} \\
=&0,
\end{aligned}$$

where the last equality is due to the fact that $E_{s+1,2} \cap \{y_{s+1} = 1\} = E_{s+1,2} \cap E_{s+1,1} = \emptyset$ conditional on $\{y_s = 1\}$. Furthermore, under Assumption A(a), we can write

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha, E_{s+1,1}) \\
=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha, E_{s+1,1}) = P(y_s = 1|w^T, y_{s-1}, \alpha, E_{s+1,1}) \\
=&P(y_s = 1|w^T, y_{s-1}, \alpha) = F_{\epsilon|\alpha}(w_s + \gamma y_{s-1} + \alpha), \qquad \text{(D.3)}
\end{aligned}$$

where the first equality uses the fact that $E_{s+1,1} \subset \{y_{s+1} = 1\}$, the second equality follows from noticing that y_s (depends only on ϵ_s) is independent of (y_{t-1}, y_{t+1}) (depend only on $(\epsilon_{s+2}, ..., \epsilon_{t+1})$) conditional on (w^T, y_{s-1}, α) and event $E_{s+1,1}$, and the third equality is because $y_s \perp E_{s+1,1}$ conditional on (w^T, y_{s-1}, α). Plugging (D.3) into (D.2) yields (A.10).

When $d_1 = 0$, (D.1) reduces to

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=&\sum_{j=2}^{3} \{P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,j}) \\
&\times P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha)\} \qquad \text{(D.4)}
\end{aligned}$$

as by definition $E_{s+1,1} \cap \{y_{s+1} = 0\} = \emptyset$.

Using Bayes' theorem and the fact that $E_{s+1,2} \cap \{y_{s+1} = 0\} = E_{s+1,2} \cap E_{s+1,3} = \emptyset$ conditional on $\{y_s = 0\}$, we have

$$\begin{aligned}
&P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,2}) \\
=&\frac{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)}{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)} \\
=&\frac{P(E_{s+1,2} \cap E_{s+1,3}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)}{P(y_{s+1} = 0, E_{s+1,2}|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha)} \\
=&0,
\end{aligned}$$

and thus

$$P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,2}) = 1. \qquad \text{(D.5)}$$

[1]The Bayes' theorem is stated mathematically as the following equation

$$P(A|B, C) = P(B|A, C)P(A|C)/P(B|C)$$

where A, B and C are events and $P(B|C) > 0$. Here, we apply Bayes' theorem by letting $A = \{y_s = 1\}$, $B = \{y_{s+1} = 1, E_{s+1,2}\}$, and $C = \{w^T, y_{s-1} = y_{t-1}, y_{t+1} = 1, \alpha\}$.

Applying similar arguments for deriving (D.3) gives

$$\begin{aligned}
&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha, E_{s+1,3}) \\
=&P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{t+1} = 0, \alpha, E_{s+1,3}) = P(y_s = 1|w^T, y_{s-1} = d, \alpha, E_{s+1,3}) \\
=&P(y_s = 1|w^T, y_{s-1}, \alpha) = F_{\epsilon|\alpha}(w_s + \gamma y_{s-1} + \alpha). \qquad \text{(D.6)}
\end{aligned}$$

Then plugging (D.5) and (D.6) into (D.4) yields (A.11). □

Proof of Lemma A.2. Again we only prove the case $\tau = s$ as the same arguments can be applied to derive the case $\tau = t$. Note that for all $j = 1, 2, 3$, we can use the law of total probability to write

$$\begin{aligned}
&P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
=&P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
&+ P(E_{s+1,j}|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
=&P(E_{s+1,j}|w^T, y_{s-1}, y_{s+1}, \alpha, y_s = 0)P(y_s = 0|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\
&+ P(E_{s+1,j}|w^T, y_{s-1}, y_{s+1}, \alpha, y_s = 1)P(y_s = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha),
\end{aligned}$$

where the second equality follows from $E_{s+1,j} \perp \{y_{t-1}, y_{t+1}\}|(w^T, y_{s-1}, y_s, y_{s+1}, \alpha)$ by Assumption A(a). Therefore, to prove (A.12)–(A.14), it suffices to verify the following equalities:

(1) $P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1} = 1, \alpha, y_s = 1) = 1$

(2) $P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1} = 1, \alpha, y_s = 0) = \frac{F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{F_{\epsilon|\alpha}(w_{s+1}+\alpha)}$

(3) $P(E_{s+1,2}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 1) = \frac{F_{\epsilon|\alpha}(w_{s+1}+\alpha)-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}$

(4) $P(E_{s+1,2}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 0) = 0$

(5) $P(E_{s+1,3}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 1) = \frac{1-F_{\epsilon|\alpha}(w_{s+1}+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}$

(6) $P(E_{s+1,3}|w^T, y_{s-1}, y_{s+1} = 0, \alpha, y_s = 0) = 1$

Equalities (1), (4), and (6) can be easily verified using the facts that $E_{s+1,1} = \{y_{s+1} = 1\}$ conditional on $\{y_s = 1\}$, $E_{s+1,2} \cap \{y_{s+1} = 0\} = \emptyset$ conditional on $\{y_s = 0\}$, and $E_{s+1,3} = \{y_{s+1} = 0\}$ conditional on $\{y_s = 0\}$, respectively.

For equality (2), note that using the conditional probability formula, we have

$$\begin{aligned}
&P(E_{s+1,1}|w^T, y_{s-1}, y_{s+1} = 1, \alpha, y_s = 0) \\
=&\frac{P(y_{s+1} = 1, E_{s+1,1}|w^T, y_{s-1}, \alpha, y_s = 0)}{P(y_{s+1} = 1|w^T, y_{s-1}, \alpha, y_s = 0)} = \frac{P(E_{s+1,1}|w^T, y_{s-1}, \alpha, y_s = 0)}{P(E_{s+1,1} \cup E_{s+1,2}|w^T, y_{s-1}, \alpha, y_s = 0)} \\
=&\frac{P(E_{s+1,1}|w_{s+1}, \alpha)}{P(E_{s+1,1} \cup E_{s+1,2}|w_{s+1}, \alpha)} = \frac{F_{\epsilon|\alpha}(w_{s+1} + \gamma + \alpha)}{F_{\epsilon|\alpha}(w_{s+1} + \alpha)}
\end{aligned}$$

where the second equality uses the fact that $\{y_{s+1}=1\}=E_{s+1,1}\cup E_{s+1,2}$ conditional on $\{y_s=0\}$, and the third equality follows by Assumption A(a).

Similar arguments, along with the fact that $\{y_{s+1}=0\}=E_{s+1,2}\cup E_{s+1,3}$ conditional on $\{y_s=1\}$, can be used to verify equalities (3) and (5). Specifically, we can write for equality (3),

$$
\begin{aligned}
&P(E_{s+1,2}|w^T,y_{s-1},y_{s+1}=0,\alpha,y_s=1)\\
=&\frac{P(y_{s+1}=0,E_{s+1,2}|w^T,y_{s-1},\alpha,y_s=1)}{P(y_{s+1}=0|w^T,y_{s-1},\alpha,y_s=1)}=\frac{P(E_{s+1,2}|w^T,y_{s-1},\alpha,y_s=1)}{P(E_{s+1,2}\cup E_{s+1,3}|w^T,y_{s-1},\alpha,y_s=1)}\\
=&\frac{P(E_{s+1,2}|w_{s+1},\alpha)}{P(E_{s+1,2}\cup E_{s+1,3}|w_{s+1},\alpha)}=\frac{F_{\epsilon|\alpha}(w_{s+1}+\alpha)-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}
\end{aligned}
$$

and analogously for equality (5),

$$
\begin{aligned}
&P(E_{s+1,3}|w^T,y_{s-1},y_{s+1},\alpha,y_s=1)\\
=&\frac{P(y_{s+1}=0,E_{s+1,3}|w^T,y_{s-1},\alpha,y_s=1)}{P(y_{s+1}=0|w^T,y_{s-1},\alpha,y_s=1)}=\frac{P(E_{s+1,3}|w^T,y_{s-1},\alpha,y_s=1)}{P(E_{s+1,2}\cup E_{s+1,3}|w^T,y_{s-1},\alpha,y_s=1)}\\
=&\frac{P(E_{s+1,3}|w_{s+1},\alpha)}{P(E_{s+1,2}\cup E_{s+1,3}|w_{s+1},\alpha)}=\frac{1-F_{\epsilon|\alpha}(w_{s+1}+\alpha)}{1-F_{\epsilon|\alpha}(w_{s+1}+\gamma+\alpha)}.
\end{aligned}
$$

Then, the proof is completed. □

We then prove Lemma A.3 with a weaker version of Assumption SD. Particularly, Assumption SD(b) will be replaced by the following stochastic dominance condition: For all $v\in\mathbb{R}$ and $d_0,d_1\in\{0,1\}$, if $w_t\geq w_s$, then

$$F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v)\geq F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v), \tag{D.7}$$

and if $w_t\leq w_s$, then

$$F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v)\leq F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1}=d_0,y_{s+1}=y_{t+1}=d_1,\alpha}(v). \tag{D.8}$$

Inequalities (D.7) and (D.8) say that, conditional on α and the same "initial" and "ending" statuses ($y_{s-1}=y_{t-1}$, $y_{s+1}=y_{t+1}$), if the value of w_t is higher than that of w_s, then w_{t+1} has a better chance of taking a large value than w_{s+1}. This restriction rules out the case in which high utility in one period has negative effects on the utility in the next period. This assumption is more likely to hold in applications where $\{w_t\}$ represents a positively autocorrelated stochastic process of the "utility", "benefits", or "profits" of a decision. This assumption is of high level, for which a sufficient, but not necessary, condition is Assumption SD(b), which is formally stated in Lemma D.1 below.

Lemma D.1. *Suppose that Assumption A is satisfied. Then inequalities (D.7) and (D.8) hold with equality, if the joint PDF of w^T conditional on α is exchangeable, i.e.,*

$$f_{w^T|\alpha}(\omega_1,...,\omega_T)=f_{w^T|\alpha}(\omega_{\pi(1)},...,\omega_{\pi(T)})$$

for all permutations $\{\pi(1),...,\pi(T)\}$ defined on the set $\mathcal{T}$.

The proof of Lemma D.1 can be found at the end of this section.

Note that inequalities (D.7) and (D.8) can be thought of as a conditional "first-order stochastic dominance" condition, which implies that, for any non-decreasing (non-increasing) function $u(\cdot)$,

$$\int u(v)dF_{w_{s+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v) \leq \int u(v)dF_{w_{t+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v)$$

$(\int u(v)dF_{w_{s+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v) \geq \int u(v)dF_{w_{t+1}|w_s=w,w_t=w',y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}(v))$ whenever $w' \geq w$. This property is needed for establishing the monotonic relation in (2.2) as demonstrated in the proof of Lemma A.3 below.

Proof of Lemma A.3. Let ϖ denote the sub-vector of w^T comprising all its elements other than w_s and w_t. Note that for all $\tau \in \{s,t\}$,

$$\begin{aligned} &P(y_\tau = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) \\ =&\int P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) dF_{\varpi|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1},\alpha}. \end{aligned} \tag{D.9}$$

In what follows, we consider two cases, $y_{s+1} = y_{t+1} = 1$ and $y_{s+1} = y_{t+1} = 0$, in turn.

Case 1 ($y_{s+1} = y_{t+1} = 1$) Plug (A.12) into (A.10) to obtain

$$\begin{aligned} &P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ =&F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)\{P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ &+ \frac{F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)}{F_{\epsilon|\alpha}(w_{\tau+1} + \alpha)}[1 - P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha)]\}. \end{aligned} \tag{D.10}$$

Let $\psi(w) \equiv F_{\epsilon|\alpha}(w + \gamma y_{\tau-1} + \alpha)$ and $\phi_1(w) \equiv F_{\epsilon|\alpha}(w + \gamma + \alpha)/F_{\epsilon|\alpha}(w + \alpha)$. Deduce from (D.10) that

$$\begin{aligned} &P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\ =&\frac{\psi(w_\tau)\phi_1(w_{\tau+1})}{1 - \psi(w_\tau) + \psi(w_\tau)\phi_1(w_{\tau+1})} \equiv G_1(w_\tau, w_{\tau+1}). \end{aligned}$$

Then, (D.9) reduces to

$$\int G_1(w_\tau, w)dF_{w_{\tau+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w),$$

and hence

$$
\begin{aligned}
&P(y_t = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\
&- P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\
=& \int G_1(w_t, w) dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w) \\
&- \int G_1(w_s, w) dF_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w) \\
=& \int \left[G_1(w_t, w) - G_1(w_s, w)\right] dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w) \\
&+ \int G_1(w_s, w) d\left[F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w) - F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=1,\alpha}(w)\right].
\end{aligned} \tag{D.11}
$$

It is easy to verify that $\psi'(\cdot) > 0$, $\phi_1'(\cdot) > 0$ (by Assumption SD(a)). Therefore, $G_1'(\cdot, w) > 0$ and $G_1'(w, \cdot) > 0$ hold true for all w. The former monotonicity result implies that the first term in (D.11) is positive if and only if $w_t \geq w_s$. The latter, together with Assumption SD(b), implies that the second term in (D.11) is positive if and only if $w_t \geq w_s$. Put these results to establish the desired result.

Case 2 $(y_{s+1} = y_{t+1} = 0)$ Plug (A.13) and (A.14) into (A.11) to obtain

$$
\begin{aligned}
&P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=& \frac{F_{\epsilon|\alpha}(w_{\tau+1} + \alpha) - F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)}{1 - F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)} P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
&+ F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)\left[\frac{1 - F_{\epsilon|\alpha}(w_{\tau+1} + \alpha)}{1 - F_{\epsilon|\alpha}(w_{\tau+1} + \gamma + \alpha)} P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha)\right. \\
&\left.+ 1 - P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha)\right].
\end{aligned} \tag{D.12}
$$

Let $\phi_0(w) \equiv [1 - F_{\epsilon|\alpha}(w + \alpha)]/[1 - F_{\epsilon|\alpha}(w + \gamma + \alpha)]$. We deduce from (D.12) that

$$
\begin{aligned}
&P(y_\tau = 1|w^T, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=& \frac{\psi(w_\tau)}{\psi(w_\tau) + \phi_0(w_{\tau+1}) - \psi(w_\tau)\phi_0(w_{\tau+1})} \equiv G_0(w_\tau, w_{\tau+1}).
\end{aligned}
$$

Then, (D.9) reduces to

$$
\int G_0(w_\tau, w) dF_{w_{\tau+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w),
$$

and hence

$$
\begin{aligned}
&P(y_t = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
&- P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 0, \alpha) \\
=& \int G_0(w_t, w) dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \\
&- \int G_0(w_s, w) dF_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \\
=& \int \left[G_0(w_t, w) - G_0(w_s, w)\right] dF_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) \\
&+ \int G_0(w_s, w) d\left[F_{w_{t+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w) - F_{w_{s+1}|w_s,w_t,y_{s-1}=y_{t-1},y_{s+1}=y_{t+1}=0,\alpha}(w)\right].
\end{aligned}
\tag{D.13}
$$

By Assumption SD(a), $\phi_0'(\cdot) < 0$. Therefore, $G_0'(\cdot, w) > 0$ and $G_0'(w, \cdot) > 0$ hold true for all w. The former monotonicity result implies that the first term in (D.13) is positive if and only if $w_t \geq w_s$. The latter, together with Assumption SD(b), implies that the second term in (D.13) is positive if and only if $w_t \geq w_s$. The proof is complete. □

Proof of Lemma A.4. The proof adopts similar arguments used in the proofs of Lemmas A.1–A.3. Here, we only outline the proof procedure and omit repetitive technical details for brevity.

First note that, under Assumptions A and SI, we can write for both $\tau = s$ and $\tau = t$,

$$
P(y_\tau = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1}, \alpha) = P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1}, \alpha). \tag{D.14}
$$

To see this, note that for $\tau = s$ and all $d_0, d_1 \in \{0, 1\}$

$$
\begin{aligned}
&P(y_s = 1|w_s, w_t, y_{s-1} = y_{t-1} = d_0, y_{s+1} = y_{t+1} = d_1, \alpha) \\
=&\frac{P(y_{t-1} = d_0, y_{t+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_s = 1, y_{s+1} = d_1, \alpha)}{P(y_{t-1} = d_0, y_{t+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha)} \\
&\times P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha) \\
=&\frac{P(y_{t-1} = d_0, y_{t+1} = d_1|w_t, y_{s+1} = d_1, \alpha)P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha)}{P(y_{t-1} = d_0, y_{t+1} = d_1|w_t, y_{s+1} = d_1, \alpha)} \\
=&P(y_s = 1|w_s, w_t, y_{s-1} = d_0, y_{s+1} = d_1, \alpha) \\
=&\frac{P(y_{s+1} = d_1|w_s, w_t, y_{s-1} = d_0, y_s = 1, \alpha)P(y_s = 1|w_s, w_t, y_{s-1} = d_0, \alpha)}{P(y_{s+1} = d_1|w_s, w_t, y_{s-1} = d_0, \alpha)} \\
=&\frac{P(y_{s+1} = d_1|w_s, y_{s-1} = d_0, y_s = 1, \alpha)P(y_s = 1|w_s, y_{s-1} = d_0, \alpha)}{P(y_{s+1} = d_1|w_s, y_{s-1} = d_0, \alpha)} \\
=&P(y_s = 1|w_s, y_{s-1} = d_0, y_{s+1} = d_1, \alpha),
\end{aligned}
$$

where the first, third, fourth, and last equalities use Bayes' theorem, and the second and fifth equalities follow by Assumptions SI(a) and A(a).[2] Using similar arguments yields the same simplification for $\tau = t$.

[2] $(y_{t-1}, y_{t+1}) \perp (w_s, y_{s-1}, y_s)|(w_t, y_{s+1}, \alpha)$ and $(y_s, y_{s+1}) \perp w_t|(w_s, y_{s-1}, \alpha)$.

For the case with $d_1 = 1$, we uses the same arguments for deriving (A.10) to write

$$\begin{aligned}
&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
=&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, E_{\tau+1,1})P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
=&F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha),
\end{aligned} \tag{D.15}$$

where the last equality follows from $E_{\tau+1,1} \subset \{y_{\tau+1} = 1\}$, Assumption SI(a), and Assumption A(a). Then, we use analogous arguments for proving Lemma A.2 to deduce

$$\begin{aligned}
&P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
=&P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, y_\tau = 1)P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
&+ P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha, y_\tau = 0)[1 - P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)] \\
=&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
&+ \frac{P(E_{\tau+1,1}|w_\tau, y_{\tau-1}, \alpha, y_\tau = 0)}{P(E_{\tau+1,1} \cup E_{\tau+1,2}|w_\tau, y_{\tau-1}, \alpha, y_\tau = 0)}[1 - P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)] \\
=&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) \\
&+ \frac{P(E_{\tau+1,1}|\alpha)}{P(E_{\tau+1,1} \cup E_{\tau+1,2}|\alpha)}[1 - P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha)],
\end{aligned} \tag{D.16}$$

where the last equality follows from Assumptions SI(a) and A(a).

Combine (D.14), (D.15) and (D.16) to solve

$$\begin{aligned}
&P(y_\tau = 1|w_s, w_t, y_{s-1} = y_{t-1}, y_{s+1} = y_{t+1} = 1, \alpha) \\
=&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 1, \alpha) = \frac{\phi_{1\alpha}\psi(w_\tau)}{1 - \psi(w_\tau) + \phi_{1\alpha}\psi(w_\tau)} \equiv \mathcal{G}_1(w_\tau),
\end{aligned}$$

where $\phi_{1\alpha} \equiv P(E_{\tau+1,1}|\alpha)/P(E_{\tau+1,1} \cup E_{\tau+1,2}|\alpha)$ is a positive constant for any given α. The monotonic relation stated in the lemma is then established by verifying the monotonicity of $\mathcal{G}_1(\cdot)$.

For the case with $d_1 = 0$, using the same arguments for deriving (A.11) yields

$$\begin{aligned}
&P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\
=&P(E_{\tau+1,2}|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\
&+ P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha, E_{\tau+1,3})P(E_{\tau+1,3}|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\
=&P(E_{\tau+1,2}|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) + F_{\epsilon|\alpha}(w_\tau + \gamma y_{\tau-1} + \alpha)P(E_{\tau+1,3}|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha),
\end{aligned} \tag{D.17}$$

where the last equality follows by $E_{\tau+1,3} \subset \{y_{\tau+1} = 0\}$, Assumption SI(a), and Assumption A(a).

Use analogous arguments for proving Lemma A.2 to obtain

$$\begin{aligned}
&P(E_{\tau+1,2}|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha) \\
=&\frac{P(E_{\tau+1,2}|\alpha)}{P(E_{\tau+1,2} \cup E_{\tau+1,3}|\alpha)}P(y_\tau = 1|w_\tau, y_{\tau-1}, y_{\tau+1} = 0, \alpha),
\end{aligned} \tag{D.18}$$

and

$$
\begin{aligned}
&P(E_{\tau+1,3}|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha)\\
=&1-P(y_\tau=1|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha)\\
&+\frac{P(E_{\tau+1,3}|\alpha)}{P(E_{\tau+1,2}\cup E_{\tau+1,3}|\alpha)}P(y_\tau=1|w_\tau, y_{\tau-1}, y_{\tau+1}=0,\alpha).
\end{aligned} \tag{D.19}
$$

Combine (D.14), (D.17), (D.18), and (D.19) to obtain

$$
\begin{aligned}
&P(y_\tau=1|w_s, w_t, y_{s-1}=y_{t-1}, y_{s+1}=y_{t+1}=1,\alpha)\\
=&P(y_\tau=1|w_\tau, y_{\tau-1}, y_{\tau+1}=1,\alpha)=\frac{\psi(w_\tau)}{\psi(w_\tau)+\phi_{0\alpha}-\phi_{0\alpha}\psi(w_\tau)}\equiv\mathcal{G}_0(w_\tau),
\end{aligned}
$$

where $\phi_{0\alpha}\equiv P(E_{\tau+1,3}|\alpha)/P(E_{\tau+1,2}\cup E_{\tau+1,3}|\alpha)$ is a positive constant for any given α. Note that $\mathcal{G}_0(w_\tau)$ is an increasing function, from which the monotonic relation stated in the lemma is established. Putting all these results together completes the proof. □

Proof of Lemma B.1. **Preparation.** Relating to the notations in Seo and Otsu (2018), $h_n=1$ (in their notations) for our estimator $\hat{\beta}$. $\xi_i(b)$ only takes value $-1, 0$, and 1, so it is bounded. Proposition 2.1 shows that β it the unique solution to $\max_{b\in\mathcal{B}}\mathbb{E}(\xi_i(b))$. The following calculation can help understand this result.

$$
\begin{aligned}
\mathbb{E}(\xi_i(b))=&\,\mathbb{E}\left\{\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|x_{i1},x_{i3}\right]\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
=&\,\mathbb{E}\left\{(\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)\right.\\
&+\mathbb{E}\left[1\left[y_{i0}=y_{i2}=y_{i4}\right](y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(\overline{y_{i0}=y_{i2}=y_{i4}}|x_{i1},x_{i3}\right))\\
&\left.\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
=&\,\mathbb{E}\left\{\mathbb{E}\left[(y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)\right.\\
&\left.\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
\equiv&\,\mathbb{E}\left\{\mathbb{E}\left[(y_{i3}-y_{i1})|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]\varphi\left(x_{i1},x_{i3}\right)\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
=&\,\mathbb{E}\left\{(\mathbb{E}\left[y_{i3}|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right]-\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2}=y_{i4},x_{i1},x_{i3}\right])\right.\\
&\left.\varphi\left(x_{i1},x_{i3}\right)\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}\\
=&\,\mathbb{E}\left\{(\mathbb{E}\left[y_{i3}|y_{i2}=y_{i4},x_{i3}\right]-\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2},x_{i1}\right])\right.\\
&\left.\varphi\left(x_{i1},x_{i3}\right)\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\},
\end{aligned}
$$

where in the second equality $\overline{A}$ denotes the complement of the set A,

$$
\varphi\left(x_{i1},x_{i3}\right)\equiv P\left(y_{i0}=y_{i2}=y_{i4}|x_{i1},x_{i3}\right)
$$

in the fourth equality, and the sixth equality follows the same argument as in the proof of Proposition 2.1.

By the stationary condition, the following is true

$$\mathbb{E}\left[y_{i3}|y_{i2}=y_{i4},x_{i3}=x\right]=\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2},x_{i1}=x\right].$$

Let

$$\phi\left(x\right)\equiv\mathbb{E}\left[y_{i3}|y_{i2}=y_{i4},x_{i3}=x\right]=\mathbb{E}\left[y_{i1}|y_{i0}=y_{i2},x_{i1}=x\right].$$

With the introduction of the above notation,

$$\mathbb{E}\left(\xi_i\left(b\right)\right)=\mathbb{E}\left\{\varphi\left(x_{i1},x_{i3}\right)\left(\phi\left(x_{i3}\right)-\phi\left(x_{i1}\right)\right)\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}. \tag{D.20}$$

From the results in the proof of Proposition 2.1, $\phi\left(x_{i3}\right)-\phi\left(x_{i1}\right)>0$ if $x_{i31}'\beta>0$, $\phi\left(x_{i3}\right)-\phi\left(x_{i1}\right)=0$ if $x_{i31}'\beta=0$, and $\phi\left(x_{i3}\right)-\phi\left(x_{i1}\right)<0$ if $x_{i31}'\beta<0$. $\varphi\left(x_{i1},x_{i3}\right)$ is a conditional probability, so $\varphi\left(x_{i1},x_{i3}\right)\geq 0$. The above observations imply that $\mathbb{E}\left(\xi_i\left(b\right)\right)$ is nonpositive and is equal to 0 if $b=\beta$. Assumption A ensures that the solution is unique. To simplify notations, let

$$\kappa\left(x_{i31}\right)\equiv\mathbb{E}\left[\varphi\left(x_{i1},x_{i3}\right)\left(\phi\left(x_{i3}\right)-\phi\left(x_{i1}\right)\right)|x_{i31}\right]. \tag{D.21}$$

It is easy to see that κ defined here is equal to the κ in the body of Lemma B.3. The above discussion implies $\kappa\left(x_{i31}\right)$ has the same sign as $x_{i31}'\beta$.

On Assumption M.i in Seo and Otsu (2018). We now try to get the derivatives of $\mathbb{E}\left(\xi_i\left(b\right)\right)$ with respect to b. $\mathbb{E}\left(\xi_i\left(b\right)\right)$ can be rewritten as

$$\mathbb{E}\left(\xi_i\left(b\right)\right)=\mathbb{E}\left\{\kappa\left(x_{i31}\right)\left(1\left[x_{i31}'b>0\right]-1\left[x_{i31}'\beta>0\right]\right)\right\}.$$

Following the same idea in Section 5 and Section 6.4 of Kim and Pollard (1990) and Section B.1 of Seo and Otsu (2018), the above expectation can be calculated using the classical differential geometry. Since the results here are obtained using essentially the same argument, we omit similar details. Define the following mapping:

$$T_b=\left(I-\left\|b\right\|_2^{-2}bb'\right)\left(I-\beta\beta'\right)+\left\|b\right\|_2^{-2}b\beta',$$

where T_b maps the region $\{x_{31}:x_{31}'b>0\}$ onto $\{x_{31}:x_{31}'\beta>0\}$, taking the boundary of $\{x_{31}:x_{31}'b>0\}$ onto the boundary of $\{x_{31}:x_{31}'\beta>0\}$. Equations (5.2) and (5.3) in Kim and Pollard (1990) imply

$$\frac{\partial}{\partial b}\mathbb{E}\left(\xi_i\left(b\right)\right)=\left\|b\right\|_2^{-2}b'\beta\left(I-\left\|b\right\|_2^{-2}bb'\right)\int 1\left[x_{31}'\beta=0\right]\kappa\left(T_bx_{31}\right)x_{31}f_{x_{31}}\left(T_bx_{31}\right)d\sigma_0,$$

where $f_{x_{31}}\left(x_{31}\right)$ is the density function of x_{i31} and σ_0 is the surface measure of the boundary of $\{x_{31}:x_{31}'\beta>0\}$.

$\frac{\partial}{\partial b}\mathbb{E}\left(\xi_i\left(b\right)\right)\big|_{b=\beta}=0$, by $T_\beta x_{31}=x_{31}$ and $1\left[x_{31}'\beta=0\right]\kappa\left(x_{31}\right)=0$. Consequently, the nonzero component of the second derivative of $\mathbb{E}\left(\xi_i\left(b\right)\right)$ only comes from the derivative of $\kappa\left(T_bx_{31}\right)$. Notice that $\frac{\partial}{\partial b}\kappa\left(T_bx_{31}\right)\big|_{b=\beta}=-\left(\frac{\partial\kappa\left(x_{31}\right)}{\partial x_{31}}'\beta\right)x_{31}$, we have

$$\left.\frac{\partial^2\mathbb{E}\left(\xi_i\left(b\right)\right)}{\partial b\partial b'}\right|_{b=\beta}=-\int 1\left[x_{31}'\beta=0\right]\left(\frac{\partial\kappa\left(x_{31}\right)}{\partial x_{31}}'\beta\right)f_{x_{31}}\left(x_{31}\right)x_{31}x_{31}'d\sigma_0.$$

Combining these results on the derivatives of $\mathbb{E}(\xi_i(b))$ implies that Assumption M.i in Seo and Otsu (2018) is satisfied with the matrix

$$V_1 \equiv -\int 1\left[x_{31}'\beta = 0\right]\left(\frac{\partial \kappa(x_{31})'}{\partial x_{31}}\beta\right) f_{x_{31}}(x_{31})\, x_{31}x_{31}' d\sigma_0. \tag{D.22}$$

By definition,

$$\left.\frac{\partial \kappa(x_{31})'}{\partial x_{31}}\beta\right|_{x_{31}'\beta=0} = \lim_{h\to 0}\left.\frac{\kappa(x_{31}+h\beta)-\kappa(x_{31})}{h}\right|_{x_{31}'\beta=0}.$$

Notice that $(x_{31}+h\beta)'\beta = h\|\beta\|$ if $x_{31}'\beta = 0$. Similar to the discussion under equation (D.20), for x_{31} satisfied with $x_{31}'\beta = 0$, $\kappa(x_{31}+h\beta) \geq 0 = \kappa(x_{31})$ if $h > 0$ and $\kappa(x_{31}+h\beta) \leq 0 = \kappa(x_{31})$ if $h < 0$. Thus, $\left.\frac{\partial \kappa(x_{31})'}{\partial x_{31}}\beta\right|_{x_{31}'\beta=0} \geq 0$, and V_1 is negative semidefinite.

On Assumption M.ii in Seo and Otsu (2018). Note

$$\xi_i(b_1) - \xi_i(b_2) = 1\left[y_{i0} = y_{i2} = y_{i4}\right](y_{i3}-y_{i1})\left(1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right)$$

and

$$(\xi_i(b_1) - \xi_i(b_2))^2 = 1\left[y_{i0} = y_{i2} = y_{i4}\right]|y_{i3}-y_{i1}|\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right|, \tag{D.23}$$

this condition can be verified by the following calculation,

$$\begin{aligned}
&\left[\mathbb{E}(\xi_i(b_1) - \xi_i(b_2))^2\right]^{1/2} \\
&= \left[\mathbb{E}\left\{\mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right|\right\}\right]^{1/2} \\
&\geq \mathbb{E}\left\{\mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right|\right\} \\
&\geq c_1\mathbb{E}\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right| \\
&\geq c_2\|b_1 - b_2\|_2,
\end{aligned}$$

where the second line holds because the value of the term in that line is smaller than 1, and a positive c_1 and c_2 can be guaranteed by Assumption A.

On Assumption M.iii in Seo and Otsu (2018). This condition can be similarly verified by

$$\begin{aligned}
&\mathbb{E}\left[\sup_{b_1,b_2\in\mathcal{B}:\|b_1-b_2\|<\varepsilon}|\xi_i(b_1) - \xi_i(b_2)|^2\right] \\
&= \mathbb{E}\left\{\sup_{b_1,b_2\in\mathcal{B}:\|b_1-b_2\|<\varepsilon}\mathbb{E}\left[|\varphi(x_{i1}, x_{i3})(\phi(x_{i3}) - \phi(x_{i1}))|\,|x_{i31}\right]\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right|\right\} \\
&\leq c_3\mathbb{E}\left\{\sup_{b_1\in\mathcal{B}:\|b_1-b_2\|<\varepsilon}\left|1\left[x_{i31}'b_1 > 0\right] - 1\left[x_{i31}'b_2 > 0\right]\right|\right\} \\
&\leq c_4\varepsilon,
\end{aligned}$$

where third line holds because φ and ϕ are conditional probability and are bounded, and the last line holds since the density of x_{31} is assumed to be bounded in Assumption 3. □

Proof of Lemma B.2. The objective function in this lemma is very similar to the one in HK. The only difference is that HK put x_{32} in the kernel $\mathcal{K}_{h_n}(\cdot)$ while we put $x_{32}'b$ and $x_{43}'b$ instead.

Seo and Otsu (2018) verified all the technical conditions needed for the estimator in HK and derived its asymptotics in Section B.1. Assumptions A and 3–5 can imply the technical conditions assumed in Section B.1 of Seo and Otsu (2018), and the conclusion follows. □

Proof of Lemma B.3. Note that

$$Z_{n,1}(\boldsymbol{s}) = n^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right) = n^{1/6} \mathbb{G}_n \left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right) + n^{2/3} \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s} n^{-1/3}\right)\right),$$

where $\mathbb{G}_n\left(\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right) = n^{-1/2}\sum_{i=1}^{n}\left[\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right) - \mathbb{E}\left(\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\right]$.

The mean of $Z_{n,1}(\boldsymbol{s})$ is $n^{2/3}\mathbb{E}\left(\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)$. With Assumptions A and 3, some calculation in the proof of Lemma B.1 yields

$$\begin{aligned} & n^{2/3}\mathbb{E}\left(\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right) \\ & = n^{2/3}\left\{\mathbb{E}\left(\xi_i(\beta)\right) + n^{-1/3} \left.\frac{\partial \mathbb{E}\left(\xi_i(b)\right)}{\partial b}\right|_{b=\beta}^{\prime} \boldsymbol{s} + \frac{1}{2} n^{-2/3} \boldsymbol{s}' \left.\frac{\partial^2 \mathbb{E}\left(\xi_i(b)\right)}{\partial b \partial b'}\right|_{b=\beta} \boldsymbol{s} + o\left(n^{-2/3}\right)\right\} \\ & = \frac{1}{2}\boldsymbol{s}' V_1 \boldsymbol{s} + o(1), \end{aligned}$$

where V_1 is defined in equation (B.1).

By definition, $H_1(\boldsymbol{s},\boldsymbol{t}) = \lim_{\alpha\to\infty}\alpha\mathbb{E}\left[\xi_i(\beta+\boldsymbol{s}/\alpha)\,\xi_i(\beta+\boldsymbol{t}/\alpha)\right]$ is the covariance kernel for the limiting distribution of $Z_{n,1}(\boldsymbol{s})$. To obtain H_1, define

$$L_1(\boldsymbol{s}-\boldsymbol{t}) \equiv \lim_{\alpha\to\infty}\alpha\mathbb{E}\left[\left(\xi_i(\beta+\boldsymbol{s}/\alpha) - \xi_i(\beta+\boldsymbol{t}/\alpha)\right)^2\right],$$
$$L_1(\boldsymbol{s}) \equiv \lim_{\alpha\to\infty}\alpha\mathbb{E}\left[\left(\xi_i(\beta+\boldsymbol{s}/\alpha) - \xi_i(\beta)\right)^2\right],$$

and

$$L_1(\boldsymbol{t}) \equiv \lim_{\alpha\to\infty}\alpha\mathbb{E}\left[\left(\xi_i(\beta+\boldsymbol{t}/\alpha) - \xi_i(\beta)\right)^2\right].$$

Notice that $\xi_i(\beta)=0$, the relationship between H_1 and L_1 is

$$H_1(\boldsymbol{s},\boldsymbol{t}) = \frac{1}{2}\left[L_1(\boldsymbol{s}) + L_1(\boldsymbol{t}) - L_1(\boldsymbol{s}-\boldsymbol{t})\right]. \tag{D.24}$$

From equations (D.20) and (D.23),

$$\begin{aligned} & \alpha\mathbb{E}\left[\left(\xi_i(\beta+\boldsymbol{s}/\alpha) - \xi_i(\beta+\boldsymbol{t}/\alpha)\right)^2\right] \\ & = \alpha\mathbb{E}\left\{\mathbb{E}\left[\left|\varphi(x_{i1},x_{i3})\left(\phi(x_{i3})-\phi(x_{i1})\right)\right| |x_{i31}\right]\left|1\left[x_{i31}'(\beta+\boldsymbol{s}/\alpha)>0\right] - 1\left[x_{i31}'(\beta+\boldsymbol{t}/\alpha)>0\right]\right|\right\} \\ & \equiv \alpha\mathbb{E}\left\{\psi(x_{i31})\left|1\left[x_{i31}'(\beta+\boldsymbol{s}/\alpha)>0\right] - 1\left[x_{i31}'(\beta+\boldsymbol{t}/\alpha)>0\right]\right|\right\}. \end{aligned}$$

where in the third line, we simplify notations by letting

$$\psi(x_{i31}) \equiv \mathbb{E}\left[\left|\varphi(x_{i1},x_{i3})\left(\phi(x_{i3})-\phi(x_{i1})\right)\right| |x_{i31}\right].$$

It is not hard to see that ψ defined here is equal to the ψ in the body of this lemma. Following Kim and Pollard (1990), we decompose x_{31} into $\varpi\beta + x_\beta$, with x_β orthogonal to β. The decomposition leads to

$$\alpha\mathbb{E}\left[\left(\xi_i\left(\beta+\boldsymbol{s}/\alpha\right)-\xi_i\left(\beta+\boldsymbol{t}/\alpha\right)\right)^2\right]$$

$$=\alpha\mathbb{E}\left\{\psi\left(x_{i31}\right)\left|1\left[x_{i31}'\left(\beta+\boldsymbol{s}/\alpha\right)>0\right]-1\left[x_{i31}'\left(\beta+\boldsymbol{t}/\alpha\right)>0\right]\right|\right\}$$

$$=\alpha\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\psi\left(\varpi\beta+x_\beta\right)\left|1\left[x_\beta'\boldsymbol{s}/\alpha+\varpi+\varpi\beta'\boldsymbol{s}/\alpha>0\right]-1\left[x_\beta'\boldsymbol{t}/\alpha+\varpi+\varpi\beta'\boldsymbol{t}/\alpha>0\right]\right|$$

$$f_{x_{31}}\left(\varpi\beta+x_\beta\right)d\varpi dx_\beta$$

$$=\alpha\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\psi\left(\varpi\beta+x_\beta\right)1\left[\frac{-x_\beta'\boldsymbol{s}/\alpha}{1+\beta'\boldsymbol{s}/\alpha}<\varpi\leq\frac{-x_\beta'\boldsymbol{t}/\alpha}{1+\beta'\boldsymbol{t}/\alpha}\right]f_{x_{31}}\left(\varpi\beta+x_\beta\right)d\varpi dx_\beta$$

$$+\alpha\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\psi\left(\varpi\beta+x_\beta\right)1\left[\frac{-x_\beta'\boldsymbol{t}/\alpha}{1+\beta'\boldsymbol{t}/\alpha}<\varpi\leq\frac{-x_\beta'\boldsymbol{s}/\alpha}{1+\beta'\boldsymbol{s}/\alpha}\right]f_{x_{31}}\left(\varpi\beta+x_\beta\right)d\varpi dx_\beta$$

$$=\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\psi\left(u/\alpha\beta+x_\beta\right)1\left[\frac{-x_\beta'\boldsymbol{s}}{1+\beta'\boldsymbol{s}/\alpha}<u\leq\frac{-x_\beta'\boldsymbol{t}}{1+\beta'\boldsymbol{t}/\alpha}\right]f_{x_{31}}\left(\left(u/\alpha\right)\beta+x_\beta\right)dudx_\beta$$

$$+\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\psi\left(u/\alpha\beta+x_\beta\right)1\left[\frac{-x_\beta'\boldsymbol{t}}{1+\beta'\boldsymbol{t}/\alpha}<u\leq\frac{-x_\beta'\boldsymbol{s}}{1+\beta'\boldsymbol{s}/\alpha}\right]f_{x_{31}}\left(\left(u/\alpha\right)\beta+x_\beta\right)dudx_\beta,$$

where the fourth equality follows by the change of variables $u=\alpha\varpi$. As $\alpha\to\infty$,

$$L_1\left(\boldsymbol{s}-\boldsymbol{t}\right)=\int_{\mathbb{R}^{K-1}}\psi\left(x_\beta\right)\left|x_\beta'\left(\boldsymbol{s}-\boldsymbol{t}\right)\right|f_{x_{31}}\left(x_\beta\right)dx_\beta,$$

Similarly,

$$L_1\left(\boldsymbol{s}\right)=\int_{\mathbb{R}^{K-1}}\psi\left(x_\beta\right)\left|x_\beta'\boldsymbol{s}\right|f_{x_{31}}\left(x_\beta\right)dx_\beta$$

and

$$L_1\left(\boldsymbol{t}\right)=\int_{\mathbb{R}^{K-1}}\psi\left(x_\beta\right)\left|x_\beta'\boldsymbol{t}\right|f_{x_{31}}\left(x_\beta\right)dx_\beta.$$

Substituting those L_1 into equation (D.24) yields

$$H_1\left(\boldsymbol{s},\boldsymbol{t}\right)=\frac{1}{2}\int_{\mathbb{R}^{K-1}}\psi\left(x_\beta\right)\left[\left|x_\beta'\boldsymbol{s}\right|+\left|x_\beta'\boldsymbol{t}\right|-\left|x_\beta'\left(\boldsymbol{s}-\boldsymbol{t}\right)\right|\right]f_{x_{31}}\left(x_\beta\right)dx_\beta.$$

□

Proof of Lemma B.4. Note

$$\hat{Z}_{n,2}\left(s\right)=\left(nh_n\right)^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)$$

$$=n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)\right)+\left(nh_n\right)^{2/3}\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)\right)$$

$$=n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)+\left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)$$

$$+n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)$$

$$+\left(nh_n\right)^{2/3}\mathbb{E}_n\left[\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right],\quad\text{(D.25)}$$

where $\mathbb{G}_n(\varsigma_{ni}(r,b)) = n^{-1/2}\sum_{i=1}^n(\varsigma_{ni}(r,b) - \mathbb{E}_n(\varsigma_{ni}(r,b)))$.

We first deal with the term in the fourth line of equation (D.25). Lemma B.2 verifies the technical conditions in Seo and Otsu (2018). Thus we can applying the result of Lemma M in Seo and Otsu (2018) on ς and it yields[3]

$$\mathbb{E}\left\{\sup_{|s|\leq C,\|b-\beta\|_2\leq Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\right\}$$
$$= n^{1/6}h_n^{1/6}\mathbb{E}\left\{\sup_{|s|\leq C,\|b-\beta\|_2\leq Mn^{-1/3}}\left|\mathbb{G}_n\left[h_n^{1/2}\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\right\}$$
$$\leq cn^{1/6}h_n^{1/6}n^{-1/6} = o(1),$$

for some positive c, any positive constants M and C. By Markov's inequality, the above yields

$$\sup_{|s|\leq C,\|b-\beta\|_2\leq Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right| = o_P(1).$$

Since $\hat{\beta}-\beta = O_P\left(n^{-1/3}\right)$, we can take M large enough so that $P\left(\left\|\hat{\beta}-\beta\right\|_2 > Mn^{-1/3}\right) < \varepsilon$ for any small $\varepsilon > 0$. For any small $\delta > 0$,

$$P\left(\sup_{|s|\leq C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\geq\delta\right)$$
$$= P\left(\left\{\sup_{|s|\leq C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\geq\delta\right\}\right.$$
$$\left.\cap\left\{\left\|\hat{\beta}-\beta\right\|_2\leq Mn^{-1/3}\right\}\right)$$
$$+ P\left(\left\{\sup_{|s|\leq C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\geq\delta\right\}\right.$$
$$\left.\cap\left\{\left\|\hat{\beta}-\beta\right\|_2 > Mn^{-1/3}\right\}\right)$$
$$\leq P\left(\sup_{|s|\leq C,\|b-\beta\|_2\leq Mn^{-1/3}} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},b\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\geq\delta\right)+\varepsilon.$$

Because the first term in the last line can be arbitrary small as $n\to\infty$, for n large enough,

$$P\left(\sup_{|s|\leq C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right|\geq\delta\right)\leq 2\varepsilon,$$

holds for any small $\delta > 0$. This implies

$$\sup_{|s|\leq C} n^{1/6}h_n^{2/3}\left|\mathbb{G}_n\left[\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)-\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\right]\right| = o_P(1). \tag{D.26}$$

For the fourth term in equation (D.25), with $\hat{\beta}-\beta = O_P\left(n^{-1/3}\right)$ and $h_n\to 0$, the expansion in equation (D.31) implies

$$(nh_n)^{2/3}\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right)\right) = (nh_n)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)+o_P(1), \tag{D.27}$$

[3] It holds by setting the δ in Lemma M of Seo and Otsu (2018) as $n^{-1/3}$.

uniformly over $|s| \leq C$. Substituting the results of equations (D.26) and (D.27) into equation (D.25) yields,

$$\begin{aligned}\hat{Z}_{n,2}(s) &= n^{1/6}h_n^{2/3}\mathbb{G}_n\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3}, \beta\right)\right) + \left(nh_n\right)^{2/3}\mathbb{E}\left(\varsigma_{ni}\left(\gamma + s\left(nh_n\right)^{-1/3}, \beta\right)\right) + o_P(1) \\ &= Z_{n,2}(s) + o_P(1),\end{aligned}$$

where the small order term holds uniformly over $|s| \leq C$ for any positive C. The claim is proved. □

Proof of Lemma B.5. We could prove the first claim in this lemma by the Taylor expansion of $\mathbb{E}\left(\varsigma_{ni}(r,\beta)\right)$ with respect to r around γ. We show a more general result instead; we derive the Taylor expansion of $\mathbb{E}\left(\varsigma_{ni}(r,b)\right)$ with respect to (r,b) around (γ,β). This more general result is useful for understanding Lemma B.5 and part of the derivation in Lemma B.4.

Recall that

$$\begin{aligned}\varsigma_{ni}(r,b) \equiv\ & \mathcal{K}_{h_n}\left(x_{i32}'b\right)\left(y_{i2}-y_{i1}\right)\left(1\left[x_{i21}'b + r\left(y_{i3}-y_{i0}\right) > 0\right] - 1\left[x_{i21}'\beta + \gamma\left(y_{i3}-y_{i0}\right) > 0\right]\right) \\ &+ \mathcal{K}_{h_n}\left(x_{i43}'b\right)\left(y_{i3}-y_{i2}\right)\left(1\left[x_{i32}'b + r\left(y_{i4}-y_{i1}\right) > 0\right] - 1\left[x_{i32}'\beta + \gamma\left(y_{i4}-y_{i1}\right) > 0\right]\right).\end{aligned}$$

To ease of notations, let

$$\begin{aligned}\vartheta_1(r,b) &\equiv \left(y_2-y_1\right)\left(1\left[x_{21}'b + r\left(y_3-y_0\right) > 0\right] - 1\left[x_{21}'\beta + \gamma\left(y_3-y_0\right) > 0\right]\right), \\ \vartheta_2(r,b) &\equiv \left(y_3-y_2\right)\left(1\left[x_{32}'b + r\left(y_4-y_1\right) > 0\right] - 1\left[x_{32}'\beta + \gamma\left(y_4-y_1\right) > 0\right]\right).\end{aligned}$$

We deal with the first component in $\varsigma_{ni}(r,b)$ first and the second term can be done analogously. First,

$$\begin{aligned}&\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{32}'b\right)\vartheta_1(r,b)\right] \\ &= \int_{\mathbb{R}^K}\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=x\right]\mathcal{K}_{h_n}\left(x'b\right)f_{x_{32}}(x)\,dx \\ &= \int_{\mathbb{R}^K}\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=x\right]\frac{1}{h_n}\mathcal{K}\left(\frac{x'b}{h_n}\right)f_{x_{32}}(x)\,dx.\end{aligned}$$

Decompose x_{32} into $x_{32} = \varpi b + x_b$, where x_b is orthogonal to b. That yields

$$\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{32}'b\right)\vartheta_1(r,b)\right] = \int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=\varpi b + x_b\right]\frac{1}{h_n}\mathcal{K}\left(\frac{\varpi}{h_n}\right)f_{x_{32}}\left(\varpi b + x_b\right)d\varpi dx_b \tag{D.28}$$

$$\begin{aligned}&= \int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=uh_nb + x_b\right]\mathcal{K}(u)\,f_{x_{32}}\left(uh_nb + x_b\right)du dx_b \\ &= \int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=x_b\right]f_{x_{32}}\left(x_b\right)dx_b \\ &+ \frac{h_n^2}{2}\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}u^2\mathcal{K}(u)\left.\frac{\partial^2\left(\mathbb{E}\left[\vartheta_1(r,b)\,|x_{32}=tb+x_b\right]f_{x_{32}}\left(tb+x_b\right)\right)}{\partial t^2}\right|_{t=t_u}du\end{aligned}$$

where in the first line we use the fact $\|b\|_2 = 1$, the second line holds by the change of variables $u = \frac{\varpi}{h_n}$, and last two lines hold by the Taylor expansion and t_u is some value between 0 and uh_n.

The bias term is of order h_n^2 by Assumption 3 and the symmetry and boundedness conditions of $\mathcal{K}$ in Assumption 4. By $nh_n^4 \to 0$ in Assumption 5, the bias term is $o\left((nh_n)^{-2/3}\right)$ and asymptotically negligible.

Similar results can be obtained for $\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{43}'b\right)\vartheta_2\left(r,b\right)\right]$.

To summarize,

$$\begin{aligned}\mathbb{E}\left(\varsigma_{ni}\left(r,b\right)\right) = &\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b \\ &+ \int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_2\left(r,b\right)|x_{43}=x_b\right] f_{x_{43}}\left(x_b\right) dx_b + o\left((nh_n)^{-2/3}\right).\end{aligned} \tag{D.29}$$

As a result, to prove the assertion in the lemma, it is enough to derive the first and second derivatives of the leading term in the above.

Notice that

$$\vartheta_1|_{(r,b)=(\gamma,\beta)} = 0.$$

Consequently, only the derivative of $E\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right]$ with respect to b in ϑ_1 will appear in

$$\frac{\partial}{\partial b}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta}.$$

That leads to

$$\begin{aligned}&\frac{\partial}{\partial b}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta} \\ =&\int_{\mathbb{R}^{K-1}} \frac{\partial}{\partial b}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]\bigg|_{(r,b)=(\gamma,\beta)} f_{x_{32}}\left(x_\beta\right) dx_\beta.\end{aligned}$$

By similar derivation as for the derivatives of $\mathbb{E}\left(\xi_i\left(b\right)\right)$, we have

$$\begin{aligned}&\frac{\partial \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]}{\partial\left(r,b'\right)'}\bigg|_{r=\gamma,b=\beta} \\ =&\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)\begin{pmatrix} y_{30} \\ x_{21}\end{pmatrix} f\left(x_{21},y_{30}|x_{32}=x_\beta\right) d\sigma_0,\end{aligned}$$

where σ_0 is the surface measure of $\{(x_{21},y_{30}) : x_{21}'\beta+\gamma y_{30}=0\}$.

$\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}'\beta=0\right)=0$ along $x_{21}'\beta+\gamma y_{30}=0$ by Proposition 2.2. Thus, the derivative above is equal to 0 and

$$\frac{\partial}{\partial\left(r,b'\right)'}\int_{\mathbb{R}^{K-1}} \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right] f_{x_{32}}\left(x_b\right) dx_b\bigg|_{r=\gamma,b=\beta} = 0.$$

The fact $\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}'\beta=0\right)=0$ along $x_{21}'\beta+\gamma y_{30}=0$ implies that only the second derivatives of $\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]$ contribute to the second derivative. By similar derivation as for the second derivative of $\mathbb{E}\left(\xi_i\left(b\right)\right)$,

$$\left.\frac{\partial^2 \mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_\beta\right]}{\partial\left(r,b'\right)'\partial\left(r,b'\right)}\right|_{r=\gamma,b=\beta}$$
$$=-\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\left(\frac{\partial \mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$
$$f\left(x_{21},y_{30}|x_{32}=x_\beta\right)\begin{pmatrix}y_{30}\\ x_{21}\end{pmatrix}\begin{pmatrix}y_{30} & x_{21}'\end{pmatrix}d\sigma_0.$$

Therefore

$$\left.\frac{\partial^2}{\partial\left(r,b'\right)'\partial\left(r,b'\right)}\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_1\left(r,b\right)|x_{32}=x_b\right]f_{x_{32}}\left(x_b\right)dx_b\right|_{r=\gamma,b=\beta}$$
$$=-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\left(\frac{\partial \mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$
$$f\left(x_{21},y_{30}|x_{32}=x_\beta\right)\begin{pmatrix}y_{30}\\ x_{21}\end{pmatrix}\begin{pmatrix}y_{30} & x_{21}'\end{pmatrix}d\sigma_0 f_{x_{32}}\left(x_\beta\right)dx_\beta$$
$$\equiv-\tilde{V}_{21}.$$

Similarly,

$$\left.\frac{\partial^2}{\partial\left(r,b'\right)'\partial\left(r,b'\right)}\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\vartheta_2\left(r,b\right)|x_{43}=x_b\right]f_{x_{43}}\left(x_b\right)dx_b\right|_{r=\gamma,b=\beta}$$
$$=-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{32}'\beta+\gamma y_{41}=0\right]\left(\frac{\partial \mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=x_\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$
$$f\left(x_{32},y_{41}|x_{43}=x_\beta\right)\begin{pmatrix}y_{41}\\ x_{32}\end{pmatrix}\begin{pmatrix}y_{41} & x_{32}'\end{pmatrix}d\sigma_0 f_{x_{43}}\left(x_\beta\right)dx_\beta$$
$$\equiv-\tilde{V}_{22}.$$

Let

$$\tilde{V}_2\equiv\tilde{V}_{21}+\tilde{V}_{22}. \tag{D.30}$$

By the Taylor expansion, Assumption 3, and equation (D.29),

$$\mathbb{E}\left(\varsigma_{ni}\left(r,b\right)\right)=-\frac{1}{2}\left(r-\gamma,\left(b-\beta\right)'\right)\tilde{V}_2\begin{pmatrix}r-\gamma\\ b-\beta\end{pmatrix}+o\left(\left\|\begin{pmatrix}r-\gamma\\ b-\beta\end{pmatrix}\right\|_2^2\right)+o\left(\left(nh_n\right)^{-2/3}\right). \tag{D.31}$$

We define V_2 as the first diagonal of $\tilde{V}_2$, that is

$$V_2\equiv e_1'\tilde{V}_2e_1,$$

where e_1 is a $(K+1)\times 1$ vector with the first element as 1 and the rest as 0. Not hard to see that

$$V_2 = -\int_{\mathbb{R}^{K-1}}\int 1\left[x_{21}'\beta+\gamma y_{30}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right) \tag{D.32}$$

$$f\left(x_{21},y_{30}|x_{32}=x_\beta\right)|y_{30}|\,d\sigma_0 f_{x_{32}}\left(x_\beta\right)dx_\beta$$

$$-\int_{\mathbb{R}^{K-1}}\int 1\left[x_{32}'\beta+\gamma y_{41}=0\right]\left(\frac{\partial\mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=x_\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\right)$$

$$f\left(x_{32},y_{41}|x_{43}=x_\beta\right)|y_{41}|\,d\sigma_0 f_{x_{43}}\left(x_\beta\right)dx_\beta.$$

Note that V_2 is a scalar.

$$\frac{\partial\mathbb{E}\left(y_{21}|x_{21},y_{30},x_{32}=x_\beta\right)'}{\partial\left(y_{30},x_{21}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\geq 0 \text{ and } \frac{\partial\mathbb{E}\left(y_{32}|x_{32},y_{41},x_{43}=\beta\right)'}{\partial\left(y_{41},x_{32}'\right)'}\begin{pmatrix}\gamma\\ \beta\end{pmatrix}\geq 0$$

hold for the same reason as in the discussion under equation (D.22). Thus, $V_2\leq 0$

Using equation (D.31),

$$\lim_{n\to\infty}\left(nh_n\right)^{2/3}\mathbb{E}_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\right)=\frac{1}{2}V_2 s^2.$$

Now, we turn to the covariance kernel. Note

$$H_2(s,t)=\lim_{n\to\infty}\left(nh_n\right)^{1/3}\mathbb{E}\left(h_n\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)\varsigma_{ni}\left(\gamma+t\left(nh_n\right)^{-1/3},\beta\right)\right).$$

Similar for the calculation of H_1 in Lemma B.1, define

$$L_2(s-t)\equiv\lim_{n\to\infty}\left(nh_n\right)^{1/3}\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)-\varsigma_{ni}\left(\gamma+t\left(nh_n\right)^{-1/3},\beta\right)\right)^2\right],$$

$$L_2(s)\equiv\lim_{n\to\infty}\left(nh_n\right)^{1/3}\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)-\varsigma_{ni}(\gamma,\beta)\right)^2\right],$$

$$L_2(t)\equiv\lim_{n\to\infty}\left(nh_n\right)^{1/3}\mathbb{E}\left[h_n\left(\varsigma_{ni}\left(\gamma+t\left(nh_n\right)^{-1/3},\beta\right)-\varsigma_{ni}(\gamma,\beta)\right)^2\right].$$

Since $\varsigma_{ni}(\gamma,\beta)=0$, $H_2(s,t)=\frac{1}{2}\left[L_2(s)+L_2(t)-L_2(s-t)\right].$

The following calculation is useful for $L_2(s-t)$.

$$\mathbb{E}\left[h_n\left(\varsigma_{ni}(r_1,\beta)-\varsigma_{ni}(r_2,\beta)\right)^2\right]$$

$$=\mathbb{E}\left\{h_n\left[\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)\left(\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right)+\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)\left(\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right)\right]^2\right\}$$

$$=\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|+h_n\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)^2\left|\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right|\right.$$

$$\left.+2h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)\left(\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right)\left(\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right)\right\}$$

$$\equiv\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1(r_1,\beta)-\vartheta_1(r_2,\beta)\right|+h_n\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)^2\left|\vartheta_2(r_1,\beta)-\vartheta_2(r_2,\beta)\right|\right\}+R_n.$$

where R_n denotes the term in the fourth line and will be shown to be asymptotic negligible.

The first term in the above can be calculated as follows,

$$\begin{aligned}&\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\right\}\\&=\int_{\mathbb{R}^K}\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=x\right]\frac{1}{h_n}\mathcal{K}\left(\frac{x'\beta}{h_n}\right)^2 f_{x_{32}}\left(x\right)dx.\end{aligned}$$

Decompose x_{32} into $x_{32}=\varpi\beta+x_\beta$, where x_β is orthogonal to β. Continue the expression in the above with this decomposition,

$$\begin{aligned}&\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{32}'\beta\right)^2\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\right\}\\&=\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=\varpi\beta+x_\beta\right]\frac{1}{h_n}\mathcal{K}\left(\frac{\varpi}{h_n}\right)^2 f_{x_{32}}\left(\varpi\beta+x_\beta\right)d\varpi dx_\beta\\&=\int_{\mathbb{R}^{K-1}}\int_{\mathbb{R}}\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=uh_n\beta+x_\beta\right]\mathcal{K}\left(u\right)^2 f_{x_{32}}\left(uh_n\beta+x_\beta\right)dudx_\beta\\&=\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=x_\beta\right]f_{x_{32}}\left(x_\beta\right)dx_\beta+O\left(h_n^2\right)\end{aligned}$$

where in the third line we substitute $u=\varpi/h_n$, in the fourth line we do Taylor expansion around $h_n=0$, the bias term is of order h_n^2 for the same reason as in equation (D.28), and $\bar{\mathcal{K}}_2=\int_{\mathbb{R}}\mathcal{K}\left(u\right)^2du$. Using Assumption 5, $\left(nh_n\right)^{2/3}h_n^2\to 0$, so the bias term is negligible. The rate of the above term can be seen from

$$\begin{aligned}&\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=x_\beta\right]\\&=\int_{\mathbb{R}}\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=\varpi,y_{30}\neq 0,x_{32}=x_\beta\right]\left|1\left[\varpi+r_1\left(y_3-y_0\right)>0\right]-1\left[\varpi+r_2\left(y_3-y_0\right)>0\right]\right|\\&P\left(y_{30}\neq 0|x_{32}=x_\beta,x_{21}'\beta=\varpi\right)f\left(x_{21}'\beta=\varpi|x_{32}=x_\beta\right)d\varpi\\&=\left|\int_{-r_1}^{-r_2}\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=\varpi,y_{30}=1,x_{32}=x_\beta\right]P\left(y_{30}=1|x_{32}=x_\beta,x_{21}'\beta=\varpi\right)f\left(x_{21}'\beta=\varpi|x_{32}=x_\beta\right)d\varpi\right|\\&+\left|\int_{r_1}^{r_2}\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=\varpi,y_{30}=-1,x_{32}=x_\beta\right]P\left(y_{30}=-1|x_{32}=x_\beta,x_{21}'\beta=\varpi\right)f\left(x_{21}'\beta=\varpi|x_{32}=x_\beta\right)d\varpi\right|\\&\propto\left|r_2-r_1\right|.\end{aligned}$$

If $r_1=\gamma+s\left(nh_n\right)^{-1/3}$ and $r_2=\gamma+t\left(nh_n\right)^{-1/3}$, $\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=x_\beta\right]\propto\left(nh_n\right)^{-1/3}$ and

$$\begin{aligned}&\lim_{n\to\infty}\left(nh_n\right)^{1/3}\mathbb{E}\left[\left|\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right|\left|x_{32}=x_\beta\right]\\&=\left\{\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]P\left(y_{30}=1|x_{32}=x_\beta,x_{21}'\beta=-\gamma\right)f\left(x_{21}'\beta=-\gamma|x_{32}=x_\beta\right)\right.\\&+\left.\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]P\left(y_{30}=-1|x_{32}=x_\beta,x_{21}'\beta=\gamma\right)f\left(x_{21}'\beta=\gamma|x_{32}=x_\beta\right)\right\}\\&\cdot\left|s-t\right|\\&=\left\{\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]f\left(y_{30}=1,x_{21}'\beta=-\gamma|x_{32}=x_\beta\right)\right.\\&\left.+\mathbb{E}\left[\left|y_{21}\right|\left|x_{21}'\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]f\left(y_{30}=-1,x_{21}'\beta=\gamma|x_{32}=x_\beta\right)\right\}\left|s-t\right|\end{aligned}$$

Therefore

$$\begin{aligned}
&\lim_{n\to\infty}(nh_n)^{1/3}\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)^2\left|\vartheta_1\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)-\vartheta_1\left(\gamma+t\left(nh_n\right)^{-1/3},\beta\right)\right|\right\}\\
&=|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]f\left(y_{30}=1,x_{21}'\beta=-\gamma|x_{32}=x_\beta\right)\right.\\
&\left.+\ \mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]f\left(y_{30}=-1,x_{21}'\beta=\gamma|x_{32}=x_\beta\right)\right\}f_{x_{32}}\left(x_\beta\right)dx_\beta.
\end{aligned}$$

For the same reason,

$$\begin{aligned}
&\lim_{n\to\infty}(nh_n)^{1/3}\mathbb{E}\left\{h_n\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)^2\left|\vartheta_2\left(\gamma+s\left(nh_n\right)^{-1/3},\beta\right)-\vartheta_2\left(\gamma+t\left(nh_n\right)^{-1/3},\beta\right)\right|\right\}\\
&=|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=-\gamma,y_{41}=1,x_{43}=x_\beta\right]f\left(y_{41}=1,x_{32}'\beta=-\gamma|x_{43}=x_\beta\right)\right.\\
&\left.+\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=\gamma,y_{41}=-1,x_{43}=x_\beta\right]f\left(y_{41}=-1,x_{32}'\beta=\gamma|x_{43}=x_\beta\right)\right\}f_{x_{43}}\left(x_\beta\right)dx_\beta.
\end{aligned}$$

Similar derivation on $R_n=2h_n\mathbb{E}\left[\mathcal{K}_{h_n}\left(x_{i32}'\beta\right)\mathcal{K}_{h_n}\left(x_{i43}'\beta\right)\left(\vartheta_1\left(r_1,\beta\right)-\vartheta_1\left(r_2,\beta\right)\right)\left(\vartheta_2\left(r_1,\beta\right)-\vartheta_2\left(r_2,\beta\right)\right)\right]$ can show that $R_n\propto(nh_n)^{-2/3}h_n$ when $r_1=\gamma+s\left(nh_n\right)^{-1/3}$ and $r_2=\gamma+t\left(nh_n\right)^{-1/3}$. So $(nh_n)^{1/3}R_n\to0$, as $n\to\infty$.

The results on $L_2\left(s-t\right)$ lead to

$$\begin{aligned}
&L_2\left(s-t\right)\\
&=|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=-\gamma,y_{30}=1,x_{32}=x_\beta\right]f\left(y_{30}=1,x_{21}'\beta=-\gamma|x_{32}=x_\beta\right)\right.\\
&\left.+\ \mathbb{E}\left[|y_{21}|\,|x_{21}'\beta=\gamma,y_{30}=-1,x_{32}=x_\beta\right]f\left(y_{30}=-1,x_{21}'\beta=\gamma|x_{32}=x_\beta\right)\right\}f_{x_{32}}\left(x_\beta\right)dx_\beta\\
&+|s-t|\,\bar{\mathcal{K}}_2\int_{\mathbb{R}^{K-1}}\left\{\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=-\gamma,y_{41}=1,x_{43}=x_\beta\right]f\left(y_{41}=1,x_{32}'\beta=-\gamma|x_{43}=x_\beta\right)\right.\\
&\left.+\mathbb{E}\left[|y_{32}|\,|x_{32}'\beta=\gamma,y_{41}=-1,x_{43}=x_\beta\right]f\left(y_{41}=-1,x_{32}'\beta=\gamma|x_{43}=x_\beta\right)\right\}f_{x_{43}}\left(x_\beta\right)dx_\beta.
\end{aligned}$$

$L_2\left(s\right)$ and $L_2\left(t\right)$ can be obtained by

$$\begin{aligned}
L_2\left(s\right)&=L_2\left(s-0\right),\\
L_2\left(t\right)&=L_2\left(t-0\right).
\end{aligned}$$

As a result

$$H_2\left(s,t\right)=\frac{1}{2}\left[L_2\left(s\right)+L_2\left(t\right)-L_2\left(s-t\right)\right],$$

which can be written as in equation (B.4). □

Proof of Lemma D.1. For the sake of brevity, we only prove the case $y_{s-1}=y_{s+1}=y_{t-1}=y_{t+1}=1$. The proofs for the other cases are similar. Denote

$$C=\left\{y_0=d_0,y_1=d_1,...,y_{s-1}=1,y_s=d_s,y_{s+1}=1,...,y_{t-1}=1,y_t=d_t,y_{t+1}=1,...,y_T=d_T\right\}$$

and $\varpi = (w_1, ...w_{s-1}, w_{s+2}, ..., w_{t-1}, w_{t+2}, ..., w_T)$. Then, by model (2.1) and Assumption A(a)

$$
\begin{aligned}
&P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)\\
=&\int P(C|\varpi, (w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) dF_{\varpi|(w_s,w_{s+1},w_t,w_{t+1})=(\omega_0,\omega_1,\omega_0',\omega_1'),\alpha}\\
=&\int p_0(w^T, \alpha)^{d_0}(1 - p_0(w^T, \alpha))^{1-d_0} \times F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha)^{d_1}(1 - F_{\epsilon|\alpha}(w_1 + \gamma d_0 + \alpha))^{1-d_1} \times \cdots\\
&\times F_{\epsilon|\alpha}(\omega_{s-1} + \gamma d_{s-2} + \alpha)F_{\epsilon|\alpha}(\omega_0 + \gamma + \alpha)^{d_s}(1 - F_{\epsilon|\alpha}(\omega_0 + \gamma + \alpha))^{1-d_s}F_{\epsilon|\alpha}(\omega_1 + \gamma d_s + \alpha) \times \cdots\\
&\times F_{\epsilon|\alpha}(\omega_{t-1} + \gamma d_{t-2} + \alpha)F_{\epsilon|\alpha}(\omega_0' + \gamma + \alpha)^{d_t}(1 - F_{\epsilon|\alpha}(\omega_0' + \gamma + \alpha))^{1-d_t}F_{\epsilon|\alpha}(\omega_1' + \gamma d_t + \alpha) \times \cdots\\
&\times F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha)^{d_T}(1 - F_{\epsilon|\alpha}(w_T + \gamma d_{T-1} + \alpha))^{1-d_T} dF_{\varpi|(w_s,w_{s+1},w_t,w_{t+1})=(\omega_0,\omega_1,\omega_0',\omega_1'),\alpha}.
\end{aligned}
$$

Given the exchangeability assumption, If $d_s = d_t$, we have

$$P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) = P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha),$$

and if $d_s \neq d_t$, we have

$$P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha) = P(\tilde{C}|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha),$$

where $\tilde{C} = \{y_0 = d_0, y_1 = d_1, ..., y_{s-1} = 1, y_s = d_t, y_{s+1} = 1, ..., y_{t-1} = 1, y_t = d_s, y_{t+1} = 1, ..., y_T = d_T\}$. Then, adding up $P(C|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)$ across all possible events C and $\tilde{C}$ yields

$$
\begin{aligned}
&P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)\\
=&P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha).
\end{aligned}
\tag{D.33}
$$

Invoke Bayes' theorem to deduce

$$
\begin{aligned}
&f_{w_s,w_{s+1},w_t,w_{t+1}|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0, \omega_1, \omega_0', \omega_1')\\
=&\frac{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0, \omega_1, \omega_0', \omega_1'), \alpha)}{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|\alpha)}\\
&\times f_{w_s,w_{s+1},w_t,w_{t+1}|\alpha}(\omega_0, \omega_1, \omega_0', \omega_1')\\
=&\frac{P(y_{s-1} = y_{t-1} = 1, y_{s+1} = y_{t+1} = 1|(w_s, w_{s+1}, w_t, w_{t+1}) = (\omega_0', \omega_1', \omega_0, \omega_1), \alpha)}{P(y_{s+1} = y_{t+1} = 1|y_{s-1} = 1, \alpha)}\\
&\times f_{w_s,w_{s+1},w_t,w_{t+1}|\alpha}(\omega_0', \omega_1', \omega_0, \omega_1)\\
=&f_{w_s,w_{s+1},w_t,w_{t+1}|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0', \omega_1', \omega_0, \omega_1),
\end{aligned}
\tag{D.34}
$$

where the second equality follows from (D.33) and the exchangeability assumption.

Applying similar arguments to obtain

$$f_{w_s,w_t|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0, \omega_0') = f_{w_s,w_t|y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_0', \omega_0). \tag{D.35}$$

Combine (D.34) and (D.35) to deduce

$$
\begin{aligned}
&f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1, \omega_1')\\
=&f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1', \omega_1).
\end{aligned}
$$

Then, the desired result follows from

$$
\begin{aligned}
& f_{w_{s+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1) \\
=& \int f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0,\omega_0'),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1,\omega_1')d\omega_1' \\
=& \int f_{w_{s+1},w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1',\omega_1)d\omega_1' \\
=& f_{w_{t+1}|(w_s,w_t)=(\omega_0',\omega_0),y_{s-1}=y_{t-1}=1,y_{s+1}=y_{t+1}=1,\alpha}(\omega_1).
\end{aligned}
$$

□

E Some Technical Details for Section 5

E.1 Numerical Bootstrap

If $\varepsilon_n = n^{-1}$, the numerical bootstrap is reduced to the classic bootstrap. Numerical bootstrap excludes the case $\varepsilon_n = n^{-1}$ and requires $n\varepsilon_n \to \infty$. The idea of numerical bootstrap is similar to the m-out-of-n bootstrap; ε_n^{-1} plays a similar role as m. As was shown in Hong and Li (2020), this procedure is less general than the m-out-of-n procedure. However, once it works, it has better finite sample performance than the m-out-of-n bootstrap. We refer to Hong and Li (2020) for the details.

Below is a heuristic illustration of why numerical bootstrap works for $\hat{\beta}$. $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right)$ can be shown to be $O_P(1)$ similarly as in Section E.3. Note that

$$
\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right) = \varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right) - \varepsilon_n^{-1/3}\left(\hat{\beta} - \beta\right) = \varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right) + o_P(1) \tag{E.1}
$$

by $n\varepsilon_n \to \infty$. Thus, the asymptotic distribution $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right)$ is the same as that of $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \beta\right)$. Let

$$
\mathcal{L}_{n,1}^*(b) \equiv n^{-1}\sum_{i=1}^n \xi_i(b) + (n\varepsilon_n)^{1/2}\cdot n^{-1}\sum_{j=1}^n\left(\xi_j^*(b) - n^{-1}\sum_{i=1}^n \xi_i(b)\right).
$$

Then $\hat{\beta}^* = \arg\max_{b\in\mathcal{B}}\mathcal{L}_{n,1}(b)$. By equation (E.1), the asymptotic distribution of $\varepsilon_n^{-1/3}\left(\hat{\beta}^* - \hat{\beta}\right)$ can be established if we can show the limiting distribution of $\varepsilon_n^{-2/3}\mathcal{L}_{n,1}^*\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)$.

The previous results suggest that

$$
\begin{aligned}
& \varepsilon_n^{-2/3}\cdot n^{-1}\sum_{i=1}^n \xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) \\
&= \varepsilon_n^{-2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right) + \varepsilon_n^{-2/3}\cdot n^{-1}\sum_{i=1}^n\left[\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right) - \mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right)\right] \\
&= \varepsilon_n^{-2/3}\mathbb{E}\left(\xi_i\left(\beta + \boldsymbol{s}\varepsilon_n^{1/3}\right)\right) + o_P(1) \\
&\xrightarrow{P} \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}
\end{aligned}
$$

over a compact set of $\boldsymbol{s}$, where the second equality holds by $n\varepsilon_n \to \infty$. The following holds by the i.i.d. sampling:

$$\varepsilon_n^{-2/3} \cdot (n\varepsilon_n)^{1/2} \cdot n^{-1} \sum_{j=1}^{n} \left(\xi_j^* \left(\beta + \boldsymbol{s}\varepsilon_n^{1/3} \right) - n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s}\varepsilon_n^{1/3} \right) \right) \rightsquigarrow W_1^* (\boldsymbol{s}),$$

where $W_1^* (\boldsymbol{s})$ is an independent copy of $W_1 (\boldsymbol{s})$. As a result,

$$\varepsilon_n^{-2/3} \mathcal{L}_{n,1}^* \left(\beta + \boldsymbol{s}\varepsilon_n^{1/3} \right) \rightsquigarrow \frac{1}{2} \boldsymbol{s}' V_1 \boldsymbol{s} + W_1^* (\boldsymbol{s}),$$

as desired.

$\hat{\gamma}$ does not directly fit into the theoretical framework of Hong and Li (2020). More specifically, condition (vi) in Theorem 4.1 in Hong and Li (2020) is not satisfied. The previous results suggest that everything in Hong and Li (2020) can go through by modifying condition (vi) to that

$$\Sigma (s,t) = \lim_{n\to\infty} (nh_n)^{1/3} \mathbb{E} \left(h_n \varsigma_{ni} \left(\gamma + s (nh_n)^{-1/3}, \beta \right) \varsigma_{ni} \left(\gamma + t (nh_n)^{-1/3}, \beta \right) \right)$$

exists for each s, t in $\mathbb{R}$. This is true by Lemma B.2. In what follows, we illustrate why numerical bootstrap works for $\hat{\gamma}$.

To concentrate on the key intuition, here we suppose that the effect of the first step estimator $\hat{\beta}$ has been handled, and it does not affect the asymptotics of $\hat{\gamma}^*$. Let

$$\mathcal{L}_{n,2}^* (r) \equiv n^{-1} \sum_{i=1}^{n} \varsigma_{ni} (r, \beta) + (n\varepsilon_n)^{1/2} \cdot n^{-1} \sum_{j=1}^{n} \left(\varsigma_{nj}^* (r, \beta) - n^{-1} \sum_{i=1}^{n} \varsigma_{ni} (r, \beta) \right),$$

where we use the same h_n in $\varsigma_{ni} (r, \beta)$ and $\varsigma_{nj}^* (r, \beta)$. The convergence rate of $\hat{\gamma}_n^*$ to γ can be shown to be $\left(\varepsilon_n^{-1} h_n \right)^{1/3}$. Thus, we only need to show the limit of $\left(\varepsilon_n^{-1} h_n \right)^{2/3} \mathcal{L}_{n,2}^* \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3} \right)$. Previous results suggest that

$$\begin{aligned}
&\left(\varepsilon_n^{-1} h_n \right)^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) \\
&= \left(\varepsilon_n^{-1} h_n \right)^{2/3} \mathbb{E} \left(\varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) \right) \\
&+ \left(\varepsilon_n^{-1} h_n \right)^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \left(\varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) - \mathbb{E} \left(\varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) \right) \right) \\
&= \left(\varepsilon_n^{-1} h_n \right)^{2/3} \mathbb{E} \left(\varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) \right) + o_P (1) \\
&\xrightarrow{P} \frac{1}{2} V_2 s^2,
\end{aligned}$$

and

$$\begin{aligned}
&\left(\varepsilon_n^{-1} h_n \right)^{2/3} \cdot (n\varepsilon_n)^{1/2} \cdot n^{-1} \sum_{j=1}^{n} \left(\varsigma_{nj}^* \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) - n^{-1} \sum_{i=1}^{n} \varsigma_{ni} \left(\gamma + s \left(\varepsilon_n^{-1} h_n \right)^{-1/3}, \beta \right) \right) \\
&\rightsquigarrow W_2^* (s) \qquad \text{(E.2)}
\end{aligned}$$

by i.i.d. and the Central Limit Theorem, where $W_2^* (s)$ is an independent copy of $W_2 (s)$. To let equation (E.2) hold, it additionally requires $\varepsilon_n^{-1} h_n \to \infty$ and $\varepsilon_n^{-1} h_n^4 \to 0$, similar to the additional restriction on m.

E.2 Classic Bootstrap

The classic bootstrap estimators for $\hat{\beta}$ and $\hat{\gamma}$, denoted as $\hat{\beta}^*$ and $\hat{\gamma}^*$, are constructed from

$$\hat{\beta}^* = \arg\max_{b\in\mathcal{B}} n^{-1}\sum_{j=1}^{n}\xi_j^*\left(b\right), \text{ and } \hat{\gamma}^* = \arg\max_{r\in\mathcal{R}} n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(r,\hat{\beta}\right).$$

Based on the proof in Abrevaya and Huang (2005), we have

$$n^{1/3}\left(\hat{\beta}^* - \beta\right) \xrightarrow{d} \arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1\left(\boldsymbol{s}\right) + W_1^*\left(\boldsymbol{s}\right)\right)$$

and

$$\left(nh_n\right)^{1/3}\left(\hat{\gamma}^* - \gamma\right) \xrightarrow{d} \arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_2s^2 + W_2\left(s\right) + W_2^*\left(s\right)\right),$$

where $W_1\left(\boldsymbol{s}\right)$ and $W_1^*\left(\boldsymbol{s}\right)$ are identical and independent Gaussian processes with zero mean and covariance kernel H_1, and $W_2\left(s\right)$ and $W_2^*\left(s\right)$ are identical and independent Gaussian processes with zero mean and covariance kernel H_2. V_1, V_2, H_1 and H_2 are the same as in Theorem 4.1.

Therefore

$$\begin{aligned} n^{1/3}\left(\hat{\beta}^* - \hat{\beta}\right) &= n^{1/3}\left(\hat{\beta}^* - \beta\right) - n^{1/3}\left(\hat{\beta} - \beta\right) \\ &\xrightarrow{d} \arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1\left(\boldsymbol{s}\right) + W_1^*\left(\boldsymbol{s}\right)\right) - \arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1\left(\boldsymbol{s}\right)\right), \end{aligned}$$

and

$$\begin{aligned} \left(nh_n\right)^{1/3}\left(\hat{\gamma}^* - \hat{\gamma}\right) &= \left(nh_n\right)^{1/3}\left(\hat{\gamma}^* - \gamma\right) - \left(nh_n\right)^{1/3}\left(\hat{\gamma} - \gamma\right) \\ &\xrightarrow{d} \arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_2s^2 + W_2\left(s\right) + W_2^*\left(s\right)\right) - \arg\max_{s\in\mathbb{R}}\left(\frac{1}{2}V_2s^2 + W_2\left(s\right)\right). \end{aligned}$$

Here, we provide a sketch showing the inconsistency of the classic bootstrap.

By similar arguments of Lemma 3 in Abrevaya and Huang (2005), the convergence rate of $\hat{\beta}^*$ to β and $\hat{\gamma}^*$ to γ can be shown be at $n^{-1/3}$ and $\left(nh_n\right)^{-1/3}$ respectively.

Define

$$Z_{n,1}^*\left(\boldsymbol{s}\right) \equiv n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\xi_j^*\left(\beta + \boldsymbol{s}n^{-1/3}\right).$$

Similar to Theorem 1 in Abrevaya and Huang (2005), one can show

$$Z_{n,1}^*\left(\boldsymbol{s}\right) \rightsquigarrow \frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s} + W_1\left(\boldsymbol{s}\right) + W_1^*\left(\boldsymbol{s}\right), \tag{E.3}$$

where $W_1(\boldsymbol{s})$ and $W_1^*(\boldsymbol{s})$ are independent and identical Gaussian processes. The intuition of this result can be seen from the following decomposition of $Z_{n,1}^*(\boldsymbol{s})$:

$$
\begin{aligned}
Z_{n,1}^*(\boldsymbol{s}) &= n^{2/3}\cdot n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)+n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right)\\
&= Z_{n,1}(\boldsymbol{s})+n^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\xi_j^*\left(\beta+\boldsymbol{s}n^{-1/3}\right)-n^{-1}\sum_{i=1}^{n}\xi_i\left(\beta+\boldsymbol{s}n^{-1/3}\right)\right),
\end{aligned}
$$

where the first term weakly converges to $\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})$, and the second term weakly converges to $W_1^*(\boldsymbol{s})$.

Since the convergence rate of $\hat{\beta}^*$ to β is $n^{-1/3}$, (E.3) implies that

$$
n^{1/3}\left(\hat{\beta}^*-\beta\right)\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})+W_1^*(\boldsymbol{s})\right),
$$

and

$$
\begin{aligned}
n^{1/3}\left(\hat{\beta}^*-\hat{\beta}\right) &= n^{-1/3}\left(\hat{\beta}^*-\beta\right)-n^{-1/3}\left(\hat{\beta}-\beta\right)\\
&\xrightarrow{d}\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})+W_1^*(\boldsymbol{s})\right)-\arg\max_{\boldsymbol{s}\in\mathbb{R}^K}\left(\frac{1}{2}\boldsymbol{s}'V_1\boldsymbol{s}+W_1(\boldsymbol{s})\right).
\end{aligned}
$$

For $\hat{\gamma}^*$, let

$$
\begin{aligned}
\hat{Z}_{n,2}^*(s) &\equiv (nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\hat{\beta}\right),\text{ and}\\
Z_{n,2}^*(s) &\equiv (nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right).
\end{aligned}
$$

The equicontinuity of $(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*(r,b)$ can be proved using similar arguments as in Theorem 1 of Abrevaya and Huang (2005). By that,

$$
\hat{Z}_{n,2}^*(s)=Z_{n,2}^*(s)+o_P(1),
$$

holds uniformly over a compact set of s. Thus we only need to establish the asymptotics of $Z_{n,2}^*(s)$. To that end, decompose $Z_{n,2}^*(s)$ as

$$
\begin{aligned}
Z_{n,2}^*(s) &= Z_{n,2}(s)+Z_{n,2}^*(s)-Z_{n,2}(s)\\
&= Z_{n,2}(s)+(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right)\\
&= Z_{n,2}(s)+(nh_n)^{2/3}\cdot n^{-1}\sum_{j=1}^{n}\left(\varsigma_{nj}^*\left(\gamma+s(nh_n)^{-1/3},\beta\right)-n^{-1}\sum_{i=1}^{n}\varsigma_{ni}\left(\gamma+s(nh_n)^{-1/3},\beta\right)\right).
\end{aligned}
$$

Using the facts that the re-sampling is i.i.d. and $n^{-1}\sum_{j=1}^{n}\varsigma_{nj}^*(r,b)$ is equicontinuous in r, it holds that

$$(nh_n)^{2/3} \cdot n^{-1} \sum_{j=1}^{n} \left(\varsigma_{nj}^{*} \left(\gamma + s (nh_n)^{-1/3}, \beta \right) - n^{-1} \sum_{i=1}^{n} \varsigma_{ni} \left(\gamma + s (nh_n)^{-1/3}, \beta \right) \right) \rightsquigarrow W_2^{*}(s),$$

where $W_2^{*}(s)$ is identically distributed as $W_2(s)$.

Lemmas B.2 and B.4 imply that

$$Z_{n,2}(s) \rightsquigarrow \frac{1}{2} V_2 s^2 + W_2(s).$$

The independence of $W_2(s)$ and $W_2^{*}(s)$ can be shown using the same arguments in the proof of Theorem 1 in Abrevaya and Huang (2005).

Combing above results implies

$$\hat{Z}_{n,2}^{*}(s) \rightsquigarrow \frac{1}{2} V_2 s^2 + W_2(s) + W_2^{*}(s).$$

Thus,

$$(nh_n)^{1/3} (\hat{\gamma}^{*} - \gamma) \xrightarrow{d} \arg\max_{s \in \mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) + W_2^{*}(s) \right),$$

and

$$\begin{aligned}(nh_n)^{1/3} (\hat{\gamma}^{*} - \hat{\gamma}) &= (nh_n)^{1/3} (\hat{\gamma}^{*} - \gamma) - (nh_n)^{1/3} (\hat{\gamma} - \gamma) \\ &\xrightarrow{d} \arg\max_{s \in \mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) + W_2^{*}(s) \right) - \arg\max_{s \in \mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) \right).\end{aligned}$$

E.3 m-out-of-n Bootstrap

Here $m \to \infty$ as $n \to \infty$, but $m/n \to 0$ as $n \to \infty$. This procedure is as follows. Draw $(y_j^{T*}, x_j^{T*\prime})'$, $j = 1, ..., m$, independently from the collection of the sample values $\left(y_1^T, x_1^{T\prime}\right)'$, $\left(y_2^T, x_2^{T\prime}\right)'$, ..., $\left(y_n^T, x_n^{T\prime}\right)'$ with replacement. Let $\hat{\beta}^{*}$ and $\hat{\gamma}^{*}$ be the estimator from the sampling observations, that is

$$\hat{\beta}^{*} = \arg\max_{b \in \mathcal{B}} m^{-1} \sum_{j=1}^{m} \xi_j^{*}(b) \text{ and } \hat{\gamma}^{*} = \arg\max_{r \in \mathcal{R}} m^{-1} \sum_{j=1}^{m} \varsigma_{nj}^{*} \left(r, \hat{\beta} \right), \tag{E.4}$$

where the bandwidth used in ς_{nj}^{*} is h_n, for simplicity. As the name suggests, this procedure only samples a small portion (m observations) from the data (n observations), with the hope of "correcting" the inconsistency of the classic bootstrap. Lee and Pun (2006) proved the consistency of m-out-of-n bootstrap for non-standard M-estimators under mild conditions. After proving the general result, they applied it to the maximum score estimator by verifying the required technical conditions. We claim that these technical conditions can be similarly verified for our estimator and

$$m^{1/3} \left(\hat{\beta}^{*} - \hat{\beta} \right) \xrightarrow{d} \arg\max_{\boldsymbol{s} \in \mathbb{R}^K} \left(\frac{1}{2} \boldsymbol{s}' V_1 \boldsymbol{s} + W_1(\boldsymbol{s}) \right)$$

and

$$(mh_n)^{1/3} (\hat{\gamma}^{*} - \hat{\gamma}) \xrightarrow{d} \arg\max_{s \in \mathbb{R}} \left(\frac{1}{2} V_2 s^2 + W_2(s) \right). \tag{E.5}$$

To make equation (E.5) hold, we additionally require $mh_n \to \infty$, $mh_n^4 \to 0$, analogous to the conditions in Assumption 5. Because of the length limitations of the paper, the details are not pursued here. Instead, we have provided a heuristic illustration.

Note $\hat{\beta}^*$ and $\hat{\gamma}^*$ in this section are obtained from expression (E.4). Let

$$Z_{m,1}^*\left(\boldsymbol{s}\right) \equiv m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \xi_j^* \left(\beta + \boldsymbol{s} m^{-1/3}\right).$$

Rewrite $Z_{m,1}^*\left(\boldsymbol{s}\right)$ as

$$\begin{aligned}
Z_{m,1}^*\left(\boldsymbol{s}\right) &= m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \left(\xi_j^* \left(\beta + \boldsymbol{s} m^{-1/3}\right) - n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right)\right) + m^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right) \\
&= m^{2/3} \cdot m^{-1} \sum_{j=1}^{m} \left(\xi_j^* \left(\beta + \boldsymbol{s} m^{-1/3}\right) - n^{-1} \sum_{i=1}^{n} \xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right)\right) \\
&\quad + m^{2/3} \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right)\right) + m^{2/3} \cdot n^{-1} \sum_{i=1}^{n} \left[\xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right) - \mathbb{E}\left(\xi_i \left(\beta + \boldsymbol{s} m^{-1/3}\right)\right)\right].
\end{aligned}$$

Intuitively, the first term in the above equation weakly converges to $W_1^*\left(\boldsymbol{s}\right)$, the second term converges to $\frac{1}{2}\boldsymbol{s}' V_1 \boldsymbol{s}$, and the last term converges to zero in probability. One can similarly show $\hat{\beta}^* - \beta = O_P\left(m^{-1/3}\right)$.

Therefore,

$$m^{1/3}\left(\hat{\beta}^* - \beta\right) \xrightarrow{d} \arg\max_{\boldsymbol{s} \in \mathbb{R}^K} \frac{1}{2}\boldsymbol{s}' V_1 \boldsymbol{s} + W_1^*\left(\boldsymbol{s}\right).$$

Finally,

$$\begin{aligned}
m^{1/3}\left(\hat{\beta}^* - \hat{\beta}\right) &= m^{1/3}\left(\hat{\beta}^* - \beta\right) - m^{1/3}\left(\hat{\beta} - \beta\right) \\
&= m^{1/3}\left(\hat{\beta}^* - \beta\right) + o_P\left(1\right) \\
&\xrightarrow{d} \arg\max_{\boldsymbol{s} \in \mathbb{R}^K} \frac{1}{2}\boldsymbol{s}' V_1 \boldsymbol{s} + W_1^*\left(\boldsymbol{s}\right).
\end{aligned}$$

Note that the distribution of $W_1^*\left(\boldsymbol{s}\right)$ is the same as that of $W_1\left(\boldsymbol{s}\right)$, and the claim is proved for $\hat{\beta}^*$. The asymptotic distribution of $\left(mh_n\right)^{1/3}\left(\hat{\gamma}^* - \hat{\gamma}\right)$ can be similarly established.

F Additional Simulation Results

F.1 Simulation Results of Designs 3–5

Table 3A: Design 3, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 2500$			$n = 5000$		
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
OY BIAS	0.7%	−0.1%	1.5%	0.5%	−0.5%	0.8%
STD	15.0%	14.9%	26.5%	11.7%	11.2%	20.8%
MAD	12.0%	11.9%	21.2%	9.3%	8.9%	16.7%
RMSE	15.0%	14.9%	26.5%	11.7%	11.2%	20.8%
HK1 BIAS	−0.7%	−0.5%	5.1%	−0.3%	−0.1%	3.6%
STD	10.2%	9.8%	21.4%	7.7%	7.9%	17.4%
MAD	8.1%	7.8%	17.6%	6.3%	6.3%	14.4%
RMSE	10.2%	9.8%	22.0%	7.7%	7.9%	17.8%
HK2 BIAS	−0.1%	−0.1%	6.3%	0.1%	0.7%	4.6%
STD	17.2%	17.1%	37.3%	14.3%	14.9%	32.2%
MAD	13.9%	13.7%	30.4%	11.6%	12.1%	26.1%
RMSE	17.2%	17.1%	37.8%	14.3%	14.9%	32.5%
	$n = 10000$			$n = 20000$		
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
OY BIAS	0.5%	0.1%	1.3%	−0.1%	0.1%	1.5%
STD	9.4%	9.6%	17.5%	7.3%	7.1%	14.0%
MAD	7.6%	7.6%	14.1%	5.9%	5.6%	11.1%
RMSE	9.4%	9.6%	17.5%	7.3%	7.1%	14.0%
HK1 BIAS	−0.4%	−0.1%	3.5%	−0.0%	−0.3%	3.6%
STD	6.3%	6.3%	13.6%	5.0%	5.0%	11.2%
MAD	5.1%	5.0%	11.3%	4.0%	4.0%	9.5%
RMSE	6.3%	6.3%	14.0%	5.0%	5.0%	11.7%
HK2 BIAS	0.4%	−0.0%	3.8%	−0.0%	0.1%	2.5%
STD	12.6%	12.2%	26.9%	11.1%	10.5%	23.5%
MAD	10.2%	9.9%	21.8%	9.0%	8.4%	18.5%
RMSE	12.6%	12.2%	27.2%	11.1%	10.5%	23.7%

Table 3B: Design 3, Numerical Bootstrap

		$n = 2500$			$n = 5000$	
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
$c = 0.8$ COV	89.7%	89.3%	86.8%	93.9%	94.5%	91.1%
LEN	106.7%	106.7%	107.5%	90.2%	90.3%	92.1%
$c = 0.9$ COV	89.4%	88.9%	86.4%	93.7%	94.4%	91.1%
LEN	104.0%	103.9%	104.5%	88.3%	88.1%	90.2%
$c = 1.0$ COV	88.9%	88.5%	85.9%	94.1%	93.9%	90.7%
LEN	101.2%	101.5%	102.1%	86.6%	86.5%	88.6%
$c = 1.1$ COV	87.6%	88.0%	85.0%	93.9%	93.9%	90.1%
LEN	99.3%	99.2%	99.8%	85.2%	85.1%	87.0%
$c = 1.2$ COV	87.6%	87.1%	84.6%	93.4%	93.7%	90.0%
LEN	97.1%	97.1%	97.5%	83.7%	83.9%	85.5%
		$n = 10000$			$n = 20000$	
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
$c = 0.8$ COV	94.3%	92.8%	91.5%	94.1%	95.0%	90.5%
LEN	75.8%	75.3%	77.6%	62.7%	62.5%	63.7%
$c = 0.9$ COV	94.2%	92.6%	90.6%	94.9%	95.0%	91.4%
LEN	74.5%	74.0%	76.9%	61.7%	61.8%	63.7%
$c = 1.0$ COV	94.2%	92.9%	90.7%	94.6%	95.4%	91.3%
LEN	73.3%	73.0%	75.9%	60.9%	60.9%	63.4%
$c = 1.1$ COV	94.0%	93.4%	91.0%	94.8%	96.1%	92.0%
LEN	72.4%	72.0%	74.9%	60.0%	60.1%	63.1%
$c = 1.2$ COV	93.8%	93.3%	91.1%	95.0%	95.6%	91.5%
LEN	71.5%	71.2%	74.1%	59.3%	59.5%	62.7%

Table 4A: Design 4, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 2500$				$n = 5000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
OY BIAS	0.3%	−0.2%	0.0%	1.7%	0.1%	−0.1%	−0.9%	0.4%
STD	16.5%	17.1%	16.8%	29.3%	13.7%	13.6%	13.8%	24.3%
MAD	13.2%	13.6%	13.4%	23.2%	10.9%	10.8%	11.1%	19.6%
RMSE	16.5%	17.1%	16.7%	29.3%	13.7%	13.6%	13.8%	24.3%
HK1 BIAS	−1.0%	−0.9%	−1.5%	10.4%	−0.5%	−1.1%	−0.5%	8.7%
STD	14.1%	14.1%	14.3%	29.5%	11.4%	11.4%	11.3%	23.5%
MAD	11.1%	11.2%	11.4%	24.9%	9.1%	9.1%	8.9%	20.2%
RMSE	14.1%	14.1%	14.3%	31.2%	11.4%	11.4%	11.3%	25.0%
HK2 BIAS	0.4%	−1.5%	−0.3%	11.2%	0.6%	−0.5%	0.1%	6.7%
STD	20.5%	21.9%	21.8%	43.6%	18.9%	18.6%	18.6%	38.7%
MAD	16.4%	17.7%	17.3%	35.9%	15.0%	14.9%	14.8%	31.2%
RMSE	20.5%	21.9%	21.8%	45.0%	18.9%	18.6%	18.6%	39.2%
	$n = 10000$				$n = 20000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
OY BIAS	−0.7%	0.6%	0.2%	2.1%	−0.3%	−0.2%	0.2%	−0.2%
STD	10.4%	10.1%	10.3%	19.8%	8.3%	8.0%	8.1%	15.5%
MAD	8.3%	8.1%	8.3%	16.1%	6.8%	6.5%	6.4%	12.4%
RMSE	10.4%	10.1%	10.3%	19.9%	8.3%	8.0%	8.1%	15.5%
HK1 BIAS	−0.3%	−0.7%	−0.1%	6.7%	0.1%	−0.8%	−0.4%	5.4%
STD	8.9%	9.4%	9.4%	18.7%	7.7%	7.6%	7.4%	15.3%
MAD	7.1%	7.6%	7.6%	15.9%	6.1%	6.1%	5.9%	12.8%
RMSE	8.9%	9.5%	9.4%	19.9%	7.7%	7.6%	7.4%	16.3%
HK2 BIAS	0.6%	0.8%	0.2%	7.6%	0.6%	−0.2%	−0.0%	5.0%
STD	16.1%	16.4%	16.5%	32.4%	14.2%	14.2%	14.1%	28.7%
MAD	13.0%	13.2%	13.1%	26.4%	11.4%	11.4%	11.3%	23.1%
RMSE	16.1%	16.4%	16.4%	33.2%	14.2%	14.2%	14.1%	29.1%

Table 4B: Design 4, Numerical Bootstrap

	$n = 2500$				$n = 5000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
$c = 0.8$ COV	89.1%	90.5%	89.3%	85.2%	93.2%	94.0%	92.1%	88.1%
LEN	111.6%	111.6%	111.7%	111.2%	95.9%	95.9%	96.0%	97.0%
$c = 0.9$ COV	88.3%	88.7%	88.7%	84.1%	92.9%	93.1%	92.6%	86.0%
LEN	108.1%	108.0%	108.0%	108.0%	93.7%	93.6%	93.6%	94.5%
$c = 1.0$ COV	86.7%	87.9%	87.9%	83.0%	92.0%	93.3%	91.5%	86.2%
LEN	104.9%	104.9%	105.0%	104.9%	91.2%	91.4%	91.2%	92.3%
$c = 1.1$ COV	85.8%	87.6%	87.2%	81.9%	92.3%	93.0%	91.5%	85.0%
LEN	102.1%	102.2%	102.3%	102.1%	89.3%	89.4%	89.3%	90.0%
$c = 1.2$ COV	86.0%	86.5%	86.4%	81.6%	91.4%	91.8%	91.0%	84.4%
LEN	99.7%	99.6%	99.6%	99.6%	87.4%	87.3%	87.6%	88.2%
	$n = 10000$				$n = 20000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
$c = 0.8$ COV	95.7%	96.0%	96.0%	88.7%	94.6%	95.9%	96.6%	92.0%
LEN	81.1%	80.9%	81.2%	83.4%	67.3%	67.4%	67.3%	70.7%
$c = 0.9$ COV	95.2%	95.8%	96.3%	88.1%	95.9%	95.7%	95.9%	92.1%
LEN	79.5%	79.6%	79.8%	81.9%	66.3%	66.4%	66.5%	69.8%
$c = 1.0$ COV	95.2%	95.7%	95.1%	88.5%	96.1%	96.0%	96.0%	91.8%
LEN	78.4%	78.3%	78.4%	80.5%	65.7%	65.7%	65.7%	68.9%
$c = 1.1$ COV	95.0%	95.6%	95.1%	87.9%	96.4%	96.5%	96.1%	91.5%
LEN	76.9%	77.2%	77.0%	79.1%	64.6%	64.8%	64.7%	68.1%
$c = 1.2$ COV	95.0%	95.4%	94.7%	86.8%	95.8%	95.5%	95.8%	91.2%
LEN	75.7%	75.9%	75.7%	77.4%	64.0%	64.1%	64.1%	66.9%

Table 5A: Design 5, Performance of $\hat{\beta}$ and $\hat{\gamma}$

	$n = 2500$					$n = 5000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
OY BIAS	−0.3%	0.4%	−0.5%	0.4%	1.9%	0.4%	−0.2%	0.3%	0.4%	0.3%
STD	18.5%	18.5%	18.1%	18.9%	36.4%	14.5%	14.1%	14.7%	14.8%	28.6%
MAD	14.8%	14.9%	14.6%	14.9%	28.9%	11.7%	11.2%	11.7%	11.7%	23.0%
RMSE	18.5%	18.5%	18.1%	18.9%	36.5%	14.5%	14.1%	14.7%	14.8%	28.5%
HK1 BIAS	−1.3%	−1.7%	−2.2%	−1.3%	12.6%	−1.7%	−0.6%	−1.1%	−1.1%	12.1%
STD	16.9%	17.5%	18.2%	18.0%	34.4%	14.4%	14.7%	14.6%	14.7%	29.0%
MAD	13.5%	14.1%	14.4%	14.2%	29.2%	11.6%	11.8%	11.8%	11.8%	25.1%
RMSE	17.0%	17.6%	18.4%	18.0%	36.6%	14.5%	14.7%	14.6%	14.8%	31.4%
HK2 BIAS	−0.5%	−0.8%	−1.5%	0.8%	14.2%	−0.7%	0.7%	−0.9%	0.3%	11.8%
STD	23.3%	24.7%	24.7%	25.2%	47.7%	22.0%	22.2%	21.8%	21.4%	42.5%
MAD	18.3%	19.7%	19.9%	20.1%	39.5%	17.8%	17.7%	17.4%	17.0%	35.7%
RMSE	23.2%	24.7%	24.7%	25.2%	49.8%	22.0%	22.2%	21.8%	21.4%	44.1%
	$n = 10000$					$n = 20000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
OY BIAS	0.6%	0.1%	−0.0%	−0.0%	0.8%	−0.0%	−0.5%	0.5%	0.1%	0.5%
STD	11.2%	11.2%	12.0%	11.5%	22.2%	9.5%	9.3%	9.3%	9.1%	18.1%
MAD	9.0%	9.1%	9.7%	9.2%	17.9%	7.5%	7.4%	7.4%	7.4%	14.5%
RMSE	11.2%	11.2%	12.0%	11.4%	22.3%	9.5%	9.3%	9.3%	9.1%	18.1%
HK1 BIAS	−0.4%	−0.7%	−0.6%	−0.9%	10.7%	−0.9%	−0.5%	−0.1%	−0.3%	9.2%
STD	12.2%	12.3%	12.5%	12.2%	24.2%	9.9%	10.1%	9.9%	10.6%	20.4%
MAD	9.8%	10.0%	9.9%	9.7%	21.2%	7.9%	8.0%	7.9%	8.4%	17.7%
RMSE	12.2%	12.3%	12.5%	12.2%	26.4%	10.0%	10.1%	9.9%	10.6%	22.4%
HK2 BIAS	0.7%	−0.2%	0.1%	1.0%	8.3%	−0.6%	0.3%	0.8%	2.1%	9.3%
STD	19.1%	19.3%	20.1%	19.4%	38.9%	17.1%	16.8%	16.6%	17.6%	34.6%
MAD	15.4%	15.6%	16.0%	15.4%	32.0%	13.7%	13.4%	13.3%	14.0%	28.6%
RMSE	19.1%	19.3%	20.1%	19.4%	39.8%	17.1%	16.8%	16.6%	17.7%	35.8%

Table 5B: Design 5, Numerical Bootstrap

	$n = 2500$					$n = 5000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
$c = 0.8$ COV	86.0%	85.0%	86.3%	84.8%	77.0%	90.0%	90.2%	91.0%	89.2%	81.9%
LEN	113.7%	113.9%	113.8%	113.7%	113.6%	99.9%	99.7%	99.8%	99.8%	100.4%
$c = 0.9$ COV	85.0%	84.2%	85.3%	83.9%	75.8%	89.5%	89.7%	90.0%	88.4%	81.5%
LEN	109.7%	109.6%	109.7%	109.7%	109.8%	96.8%	96.5%	96.9%	96.8%	97.3%
$c = 1.0$ COV	83.6%	83.2%	85.0%	83.1%	74.8%	89.1%	89.2%	89.0%	87.7%	80.2%
LEN	106.3%	106.2%	106.3%	106.2%	106.6%	93.9%	93.9%	93.9%	93.9%	94.6%
$c = 1.1$ COV	83.3%	83.1%	84.2%	82.5%	73.9%	87.4%	88.8%	88.4%	87.4%	79.3%
LEN	103.2%	103.2%	103.3%	103.3%	103.4%	91.6%	91.6%	91.7%	91.6%	92.2%
$c = 1.2$ COV	82.2%	82.0%	83.9%	81.7%	73.1%	86.6%	87.4%	87.9%	86.6%	77.8%
LEN	100.3%	100.3%	100.4%	100.4%	100.7%	89.2%	89.3%	89.1%	89.2%	90.0%
	$n = 10000$					$n = 20000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
$c = 0.8$ COV	94.6%	95.2%	94.6%	94.9%	86.9%	96.0%	97.3%	97.0%	97.2%	90.0%
LEN	85.7%	85.8%	85.8%	85.9%	88.0%	71.8%	71.9%	72.0%	71.9%	75.8%
$c = 0.9$ COV	94.3%	94.6%	94.2%	94.5%	86.7%	97.0%	97.1%	97.2%	96.4%	88.8%
LEN	83.7%	83.7%	83.8%	83.6%	85.6%	70.7%	70.7%	70.9%	70.7%	74.4%
$c = 1.0$ COV	93.8%	93.5%	93.9%	94.4%	86.2%	96.3%	96.7%	97.2%	96.8%	89.3%
LEN	81.9%	81.9%	81.9%	81.8%	83.6%	69.6%	69.5%	69.7%	69.6%	72.9%
$c = 1.1$ COV	93.2%	92.8%	93.6%	93.8%	86.1%	95.6%	96.8%	96.7%	96.5%	88.7%
LEN	80.1%	80.1%	80.0%	80.0%	81.6%	68.6%	68.4%	68.6%	68.3%	71.3%
$c = 1.2$ COV	92.7%	92.4%	93.2%	93.1%	84.8%	95.6%	96.4%	96.1%	96.3%	88.8%
LEN	78.4%	78.3%	78.4%	78.4%	79.8%	67.6%	67.5%	67.4%	67.4%	70.2%

F.2 Additional Monte Carlo Experiments: Designs 6–8

To investigate the impact of serial correlations of x_{it} on our estimator and the proposed inference procedure, we conduct additional simulations for several designs. These designs are the same as in Designs 3–5, except that we allow for x to be auto-correlated, similar to Design 2, as opposed to Design 1.

We consider Designs 6–8, which employ the same models as those in Designs 3–5, respectively, but with serially dependent x_{it}. More specifically:

$$\begin{aligned} x_{i0,j} &= \frac{\sqrt{15}}{4} u_{it,j} + \frac{1}{4} u_{it,k+1}, \; j = 1, 2, ..., k \text{ and} \\ x_{it,j} &= 0.25 x_{it-1,j} + \sqrt{1 - 0.25^2} \left(\frac{\sqrt{15}}{4} u_{it,j} + \frac{1}{4} u_{it,k+1} \right), j = 1, 2, ..., k \text{ for all } t \geq 1, \end{aligned}$$

with $(u_{it,1}, u_{it,2}, ..., u_{it,k+1})$ distributed as $N\left(0_{(k+1)\times 1}, I_{(k+1)\times(k+1)}\right)$, and $(u_{it,1}, u_{it,2}, ..., u_{it,k+1})$ being i.i.d. across i and t. The parameter k is set to be 3, 4, or 5 in Designs 6 through 8, respectively.

To conserve space, we only report the inference results for $c = 1$. We report the BIAS, STD, MAD, RMSE, COV, and LEN for our estimators. All results are collected into one table for each design. The tables are numbered corresponding to the names of the designs. For instance, results for Design 6 are presented in Table 6.

A brief summary of our findings is as follows: Our estimation results seem to remain relatively unchanged with serially correlated x. Importantly, they are not significantly biased. The inference procedure performs well, maintaining the same level of performance as designs with serially independent x_{it}.

Table 6: Design 6

	$n = 2500$			$n = 5000$		
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
OY BIAS	−0.8%	0.5%	2.5%	0.8%	−0.4%	2.1%
STD	15.4%	15.1%	25.2%	11.2%	11.6%	20.3%
MAD	12.4%	12.2%	20.0%	9.0%	9.4%	16.3%
RMSE	15.4%	15.1%	25.3%	11.2%	11.6%	20.4%
COV	90.7%	89.8%	87.4%	93.0%	92.5%	90.7%
LEN	100.9%	101.3%	100.6%	86.5%	86.3%	87.1%
	$n = 10000$			$n = 20000$		
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\gamma}$
OY BIAS	0.4%	−0.5%	2.8%	−0.2%	−0.1%	1.1%
STD	9.4%	9.9%	16.5%	7.6%	7.6%	13.2%
MAD	7.4%	8.0%	13.4%	6.1%	6.1%	10.5%
RMSE	9.4%	9.9%	16.7%	7.6%	7.6%	13.2%
COV	95.7%	92.9%	92.7%	93.3%	93.6%	92.1%
LEN	73.0%	72.7%	74.0%	60.6%	60.7%	61.2%

Table 7: Design 7

	$n = 2500$				$n = 5000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
OY BIAS	0.2%	−0.6%	0.6%	3.1%	−0.5%	−0.1%	0.3%	1.7%
STD	16.1%	16.3%	16.9%	29.8%	13.4%	13.6%	13.5%	23.5%
MAD	13.2%	13.1%	13.5%	23.9%	10.7%	10.9%	10.9%	19.1%
RMSE	16.1%	16.3%	16.9%	29.9%	13.4%	13.6%	13.5%	23.6%
COV	87.0%	87.7%	86.7%	80.5%	90.8%	90.5%	90.9%	85.8%
LEN	104.7%	104.8%	104.7%	104.2%	91.3%	91.3%	91.3%	91.4%
	$n = 10000$				$n = 20000$			
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\gamma}$
OY BIAS	−0.1%	−0.1%	−0.1%	2.8%	−0.5%	0.5%	0.1%	2.1%
STD	10.9%	10.7%	10.6%	19.2%	8.4%	8.2%	8.5%	15.0%
MAD	8.8%	8.5%	8.4%	15.6%	6.8%	6.6%	6.9%	12.0%
RMSE	10.9%	10.7%	10.6%	19.4%	8.5%	8.3%	8.5%	15.1%
COV	94.3%	93.5%	94.4%	89.8%	95.7%	95.9%	96.1%	91.9%
LEN	78.5%	78.1%	78.2%	79.0%	65.4%	65.7%	65.6%	67.3%

Table 8: Design 8

	$n = 2500$					$n = 5000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
OY BIAS	−1.7%	1.1%	0.1%	−0.1%	6.4%	0.1%	−0.0%	0.3%	−0.2%	2.2%
STD	18.8%	18.6%	18.0%	18.1%	32.8%	14.7%	15.2%	14.6%	15.1%	27.3%
MAD	15.0%	15.2%	14.2%	14.5%	26.4%	11.8%	12.1%	11.6%	12.2%	22.2%
RMSE	18.8%	18.6%	18.0%	18.1%	33.4%	14.7%	15.2%	14.6%	15.1%	27.4%
COV	84.1%	82.5%	83.3%	83.8%	78.1%	88.5%	89.4%	89.6%	89.2%	83.0%
LEN	106.2%	106.2%	106.0%	106.2%	106.0%	93.6%	93.7%	93.7%	93.7%	94.0%
	$n = 10000$					$n = 20000$				
	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\gamma}$
OY BIAS	0.3%	−0.2%	0.3%	−0.2%	2.9%	−0.2%	−0.1%	−0.2%	0.6%	1.3%
STD	11.5%	11.7%	11.5%	11.8%	21.1%	9.4%	9.3%	8.9%	9.0%	16.2%
MAD	9.2%	9.3%	9.1%	9.4%	16.9%	7.5%	7.4%	7.2%	7.2%	13.1%
RMSE	11.5%	11.6%	11.5%	11.8%	21.3%	9.4%	9.3%	8.9%	9.0%	16.3%
COV	93.4%	94.6%	93.8%	92.3%	87.4%	95.7%	96.3%	96.8%	96.4%	92.2%
LEN	81.8%	81.8%	81.6%	81.7%	82.5%	69.6%	69.5%	69.5%	69.5%	71.8%

References

Abrevaya, J. and J. Huang (2005): "On the bootstrap of the maximum score estimator," *Econometrica*, 73, 1175–1204.

Hong, H. and J. Li (2020): "The numerical bootstrap," *The Annals of Statistics*, 48, 397–412.

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

Lee, S. M. S. and M. C. Pun (2006): "On m out of n bootstrapping for nonstandard m-estimation with nuisance parameters," *Journal of American Statistical Association*, 101, 1185–1197.

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