Supplemental to "Bootstrap-Based Inference for Cube Root Asymptotics"*

Matias D. Cattaneo† Michael Jansson‡ Kenichi Nagasawa§

June 3, 2020

Abstract

This supplemental appendix contains proofs and other theoretical results that may be of independent interest. It also offers more details on the examples and simulation evidence presented in the paper.

*Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1459931 and SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through grants SES-1459967 and SES-1947662 and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78).

†Department of Operations Research and Financial Engineering, Princeton University.

‡Department of Economics, University of California at Berkeley and CREATES.

§Department of Economics, University of Warwick.

Contents

A.1 Proofs of Main Results

A.1.1 Proof of Theorem 1

As explained in the paper, Theorem 1 follows from ten technical lemmas. The remainder of this subsection presents those lemmas and their proofs.

The first lemma can be used to show that $\hat{\boldsymbol{\theta}}_n$ is consistent.

Lemma A.1 *Suppose Condition CRA(i) holds. Then* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_{\mathbb{P}}(1)$ *if*

$$\hat{M}_n(\hat{\boldsymbol{\theta}}_n) \geq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \hat{M}_n(\boldsymbol{\theta}) - o_{\mathbb{P}}(1).$$

Proof of Lemma A.1. It suffices to show that every $\delta > 0$ admits a constant $c_\delta > 0$ such that

$$\mathbb{P}\left[\hat{M}_n(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} \hat{M}_n(\boldsymbol{\theta}) > c_\delta\right] \to 1. \tag{A.1}$$

By assumption, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})| = o(1)$. Also, by Pollard (1989, Theorem 4.2),

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{M}_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta})| = O_{\mathbb{P}}\left(\sqrt{\frac{\mathbb{E}[\bar{m}_n(\mathbf{z})^2]}{n}}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nq_n}}\right) = o_{\mathbb{P}}(1).$$

As a consequence, for any $\delta > 0$,

$$\hat{M}_n(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} \hat{M}_n(\boldsymbol{\theta}) = M_0(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) + o_{\mathbb{P}}(1),$$

so (A.1) is satisfied with $c_\delta = [M_0(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta})]/2 > 0$. ∎

Assuming the derivatives exist, let $\dot{M}_n(\boldsymbol{\theta})$ and $\ddot{M}_n(\boldsymbol{\theta})$ denote $\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$ and $\partial^2 M_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, respectively. If M_n is twice continuously differentiable on a neighborhood $\boldsymbol{\Theta}_n$ of $\boldsymbol{\theta}_0$, then it follows from Taylor's theorem that

$$\left|M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbf{H}_n(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right| \leq \dot{C}_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| + \frac{1}{2}\ddot{C}_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||^2, \tag{A.2}$$

for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}_n$, where $\mathbf{H}_n = -\ddot{M}_n(\boldsymbol{\theta}_0)$, $\dot{C}_n = ||\dot{M}_n(\boldsymbol{\theta}_0)||$, and $\ddot{C}_n = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_n} ||\ddot{M}_n(\boldsymbol{\theta}) - \ddot{M}_n(\boldsymbol{\theta}_0)||$.

As an immediate consequence of (A.2), we have the following convergence result about Q_n.

Lemma A.2 *Suppose Condition CRA(ii) holds. Then* Q_n *converges compactly to* $\mathcal{Q}_0$; *that is,*

$$\sup_{||\mathbf{s}||\leq K} |Q_n(\mathbf{s}) - \mathcal{Q}_0(\mathbf{s})| \to 0$$

for any $K > 0$.

Proof of Lemma A.2. Let $K > 0$ be given and suppose n is large enough that $Kr_n^{-1} \leq \delta$, where $\delta > 0$ is as in Condition CRA(ii). Using (A.2) with $\boldsymbol{\Theta}_n = \boldsymbol{\Theta}_0^{Kr_n^{-1}}$, we have

$$\begin{aligned} |Q_n(\mathbf{s}) - \mathcal{Q}_0(\mathbf{s})| &= \left|r_n^2[M_n(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) - M_n(\boldsymbol{\theta}_0)] + \frac{1}{2}\mathbf{s}'\mathbf{H}_0\mathbf{s}\right| \\ &\leq \frac{1}{2}\left|\mathbf{s}'(\mathbf{H}_n - \mathbf{H}_0)\mathbf{s}\right| + r_n\dot{C}_n||\mathbf{s}|| + \frac{1}{2}\ddot{C}_n||\mathbf{s}||^2 = (K + K^2)o(1) \end{aligned}$$

uniformly in $\mathbf{s}$ with $||\mathbf{s}|| \leq K$, where the last equality uses $r_n \dot{C}_n = r_n ||\dot{M}_n(\boldsymbol{\theta}_0)|| \to 0$ along with the facts that

$$\mathbf{H}_n - \mathbf{H}_0 = -[\ddot{M}_n(\boldsymbol{\theta}_0) - \ddot{M}_0(\boldsymbol{\theta}_0)] \to 0, \qquad \ddot{M}_0(\boldsymbol{\theta}_0) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} M_0(\boldsymbol{\theta}),$$

and

$$\begin{aligned} \ddot{C}_n &= \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{K r_n^{-1}}} ||\ddot{M}_n(\boldsymbol{\theta}) - \ddot{M}_n(\boldsymbol{\theta}_0)|| \\ &\leq 2 \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{K r_n^{-1}}} ||\ddot{M}_n(\boldsymbol{\theta}) - \ddot{M}_0(\boldsymbol{\theta})|| + \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{K r_n^{-1}}} ||\ddot{M}_0(\boldsymbol{\theta}) - \ddot{M}_0(\boldsymbol{\theta}_0)|| \to 0. \qquad \blacksquare \end{aligned}$$

The next lemma can be used to obtain the rate of convergence of $\hat{\boldsymbol{\theta}}_n$.

Lemma A.3 *Suppose Conditions CRA(ii)-(iii) hold. Then* $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ *if* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_{\mathbb{P}}(1)$ *and if*

$$\hat{M}_n(\hat{\boldsymbol{\theta}}_n) \geq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \hat{M}_n(\boldsymbol{\theta}) - o_{\mathbb{P}}(r_n^{-2}).$$

Proof of Lemma A.3. For any $\delta > 0$ and any $K \in \mathbb{N}$, $\mathbb{P}[r_n ||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| > 2^K]$ is no greater than

$$\begin{aligned} &\mathbb{P}[\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n) \geq \delta r_n^{-2}] + \mathbb{P}[||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| > \delta/2] \\ &+ \sum_{j \geq K, 2^j \leq \delta r_n} \mathbb{P}\left[\sup_{2^{j-1} < r_n ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq 2^j} \hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) \geq -\delta r_n^{-2} \right]. \end{aligned}$$

By assumption, the probabilities on the first line go to zero for any $\delta > 0$. As a consequence, it suffices to show that the sum on the last line can be made arbitrarily small (for large n) by making $\delta > 0$ small and K large.

To do so, let $\delta > 0$ be small enough so that Conditions CRA(ii)-(iii) are satisfied and

$$c(\delta) = \liminf_{n \to \infty} \frac{1}{16} [\lambda_{\min}(\mathbf{H}_n) - \ddot{C}_n^\delta] > 0,$$

where $\ddot{C}_n^\delta = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\delta} ||\ddot{M}_n(\boldsymbol{\theta}) - \ddot{M}_n(\boldsymbol{\theta}_0)||$ and where $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of the argument. Then, for all n large enough and for any pair $(j, K)' \in \mathbb{N}^2$ with $j \geq K$, we have

$$M_n(\boldsymbol{\theta}_0) - \sup_{2^{j-1} < r_n ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq 2^j} M_n(\boldsymbol{\theta}) - \delta r_n^{-2} \geq 2^{2j} c_{n,K}(\delta) r_n^{-2},$$

where $c_{n,K}(\delta) = [\lambda_{\min}(\mathbf{H}_n) - \ddot{C}_n^\delta]/8 - 2^{-K} r_n \dot{C}_n - 2^{-2K}\delta$ and where the inequality uses the following implication of (A.2): If $\lambda_{\min}(\mathbf{H}_n) - \ddot{C}_n^\delta \geq 0$ and if $\boldsymbol{\Theta}_n'$ is a subset of $\boldsymbol{\Theta}_n$, then

$$M_n(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_n'} M_n(\boldsymbol{\theta}) \geq \frac{1}{2} [\lambda_{\min}(\mathbf{H}_n) - \ddot{C}_n^\delta] \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_n'} ||\boldsymbol{\theta} - \boldsymbol{\theta}_0||^2 - \dot{C}_n \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_n'} ||\boldsymbol{\theta} - \boldsymbol{\theta}_0||.$$

Choosing n and K large enough, we may assume that $c_{n,K}(\delta) \geq c(\delta)$, in which case

$$\begin{aligned}
&\sum_{j\geq K, 2^j\leq \delta r_n} \mathbb{P}\left[\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\leq 2^j} \hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) \geq -\delta r_n^{-2}\right] \\
\leq & \sum_{j\geq K, 2^j\leq \delta r_n} \mathbb{P}\left[\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\leq 2^j} \{\hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}_0)\} \geq 2^{2j} c_{n,K}(\delta) r_n^{-2}\right] \\
\leq & \sum_{j\geq K, 2^j\leq \delta r_n} \mathbb{P}\left[\sup_{r_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\leq 2^j} ||\hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}_0)|| \geq 2^{2j} c(\delta) r_n^{-2}\right] \\
\leq & \frac{r_n^2}{c(\delta)} \sum_{j\geq K, 2^j\leq \delta r_n} 2^{-2j} \mathbb{E}\left[\sup_{r_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\leq 2^j} ||\hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}_0)||\right],
\end{aligned}$$

where the last inequality uses the Markov inequality.

Under Condition CRA(iii), $q_n \sup_{0\leq\delta'\leq\delta} \mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$ and it follows from Pollard (1989, Theorem 4.2) that the sum on the last line is bounded by a constant multiple of

$$r_n^2 \sum_{j\geq K, 2^j\leq \delta r_n} 2^{-2j} \sqrt{\frac{\mathbb{E}[\bar{d}_n^{2^j/r_n}(\mathbf{z})^2]}{n}} \leq \sqrt{q_n \sup_{0\leq\delta'\leq\delta} \mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta']} \sum_{j\geq K} 2^{-3j/2},$$

which can be made arbitrarily small by making K large. ∎

In combination, the next two lemmas can be used to show that $\hat{G}_n \rightsquigarrow \mathcal{G}_0$ in the topology of uniform convergence on compacta.

Lemma A.4 *Suppose Conditions CRA(iii)-(iv) hold and suppose $Q_n(\mathbf{s}) = o(\sqrt{n})$ for every $\mathbf{s} \in \mathbb{R}^d$. Then $\hat{G}_n$ converges to $\mathcal{G}_0$ in the sense of weak convergence of finite-dimensional projections.*

Proof of Lemma A.4. Because $\hat{G}_n(\mathbf{s}) = n^{-1/2}\sum_{i=1}^n \psi_n(\mathbf{z}_i;\mathbf{s})$, where

$$\psi_n(\mathbf{z};\mathbf{s}) = \sqrt{r_n q_n}[m_n(\mathbf{z},\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) - m_n(\mathbf{z},\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) + M_n(\boldsymbol{\theta}_0)]\mathbb{1}(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1} \in \boldsymbol{\Theta})$$

the result follows from the Cramér-Wold device if

$$\mathbb{E}[\psi_n(\mathbf{z};\mathbf{s})\psi_n(\mathbf{z};\mathbf{t})] \to \mathcal{C}_0(\mathbf{s},\mathbf{t}) \qquad \forall \mathbf{s},\mathbf{t} \in \mathbb{R}^d,$$

and if the following Lyapunov condition is satisfied:

$$\frac{1}{n}\mathbb{E}[\psi_n(\mathbf{z};\mathbf{s})^4] \to 0 \qquad \forall \mathbf{s} \in \mathbb{R}^d.$$

Let $\mathbf{s},\mathbf{t} \in \mathbb{R}^d$ be given and suppose without loss of generality that $\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}, \boldsymbol{\theta}_0 + \mathbf{t}r_n^{-1} \in \boldsymbol{\Theta}$. Then, using $Q_n(\mathbf{s}) = o(\sqrt{n})$ and the representation

$$\psi_n(\mathbf{z};\mathbf{s}) = \sqrt{r_n q_n}[m_n(\mathbf{z},\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) - m_n(\mathbf{z},\boldsymbol{\theta}_0)] - \frac{1}{\sqrt{n}}Q_n(\mathbf{s}),$$

we have

$$\begin{aligned}
& \mathbb{E}[\psi_n(\mathbf{z};\mathbf{s})\psi_n(\mathbf{z};\mathbf{t})] \\
= \; & r_n q_n \mathbb{E}[\{m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}r_n^{-1})-m_n(\mathbf{z},\boldsymbol{\theta}_0)\}\{m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{t}r_n^{-1})-m_n(\mathbf{z},\boldsymbol{\theta}_0)\}]-\frac{1}{n}Q_n(\mathbf{s})Q_n(\mathbf{t}) \\
\to \; & \mathcal{C}_0(\mathbf{s},\mathbf{t})
\end{aligned}$$

and, using $\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4]=o(q_n^{-3}r_n)$ (for $\delta_n=O(r_n^{-1})$),

$$\frac{1}{16n}\mathbb{E}[\psi(\mathbf{z};\mathbf{s})^4]\le\frac{r_n^2q_n^2}{n}\mathbb{E}[|m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}r_n^{-1})-m_n(\mathbf{z},\boldsymbol{\theta}_0)|^4]+\frac{1}{n^3}Q_n(\mathbf{s})^4=o\left(\frac{r_n^3}{nq_n}+\frac{1}{n}\right)=o(1),$$

as was to be shown. ∎

Lemma A.5 *Suppose Conditions CRA(iii) and CRA(v) hold. Then* $\{\hat{G}_n(\mathbf{s}):||\mathbf{s}||\le K\}$ *is stochastically equicontinuous for every* $K>0$*; that is,*

$$\sup_{\substack{||\mathbf{s}-\mathbf{t}||\le\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\le K}}|\hat{G}_n(\mathbf{s})-\hat{G}_n(\mathbf{t})|\to_{\mathbb{P}}0$$

for any $K>0$ *and for any* $\Delta_n>0$ *with* $\Delta_n=o(1)$.

Proof of Lemma A.5. Let $K>0$ be given. As in the proof of Kim and Pollard (1990, Lemma 4.6) and using the fact that $q_n\delta_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^2]=O(1)$ (for $\delta_n=O(r_n^{-1})$), it suffices to show that

$$r_n\sup_{\substack{||\mathbf{s}-\mathbf{t}||\le\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\le K}}\frac{q_n}{n}\sum_{i=1}^n d_n(\mathbf{z}_i;\mathbf{s},\mathbf{t})^2\to_{\mathbb{P}}0,$$

where $d_n(\mathbf{z};\mathbf{s},\mathbf{t})=|m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}r_n^{-1})-m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{t}r_n^{-1})|/2$.

For any $C>0$ and any $\mathbf{s},\mathbf{t}\in\mathbb{R}^d$ with $||\mathbf{s}||,||\mathbf{t}||\le K$,

$$\begin{aligned}
\frac{q_n}{n}\sum_{i=1}^n d_n(\mathbf{z}_i;\mathbf{s},\mathbf{t})^2 \le \; & \frac{q_n}{n}\sum_{i=1}^n \bar{d}_n^{Kr_n^{-1}}(\mathbf{z}_i)^2\mathbb{1}(q_n\bar{d}_n^{Kr_n^{-1}}(\mathbf{z}_i)>C) \\
& +C\mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})] \\
& +C\frac{1}{n}\sum_{i=1}^n\{d_n(\mathbf{z}_i;\mathbf{s},\mathbf{t})-\mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]\},
\end{aligned}$$

and therefore

$$\begin{aligned}
r_n\mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}||\le\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\le K}}\frac{q_n}{n}\sum_{i=1}^n d_n(\mathbf{z};\mathbf{s},\mathbf{t})^2\right] \le \; & q_nr_n\mathbb{E}\left[\bar{d}_n^{Kr_n^{-1}}(\mathbf{z})^2\mathbb{1}(q_n\bar{d}_n^{Kr_n^{-1}}(\mathbf{z})>C)\right] \\
& +Cr_n\sup_{\substack{||\mathbf{s}-\mathbf{t}||\le\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\le K}}\mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})] \\
& +Cr_n\mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}||\le\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\le K}}|\frac{1}{n}\sum_{i=1}^n\{d_n(\mathbf{z}_i;\mathbf{s},\mathbf{t})-\mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]\}|\right].
\end{aligned}$$

For large n, the first term on the majorant side can be made arbitrarily small by making C

large. Also, for any fixed C, the second term tends to zero because $\Delta_n \to 0$. Finally, Pollard (1989, Theorem 4.2) can be used to show that for fixed C and for large n, the last term is bounded by a constant multiple of

$$r_n \sqrt{\frac{\mathbb{E}[\bar{d}_n^{Kr_n^{-1}}(\mathbf{z})^2]}{n}} = \frac{\sqrt{K}}{r_n} \sqrt{q_n \mathbb{E}[\bar{d}_n^{Kr_n^{-1}}(\mathbf{z})^2/(Kr_n^{-1})]} = O\left(\frac{1}{r_n}\right) = o(1). \blacksquare$$

The analysis of $\tilde{\boldsymbol{\theta}}_n^*$ also relies on five lemmas, each of which is a natural bootstrap analog of a lemma used to analyze $\hat{\boldsymbol{\theta}}_n$. The following lemma can be used to show that $\tilde{\boldsymbol{\theta}}_n^*$ is consistent in the sense that $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(1)$.

Lemma A.6 *Suppose Condition CRA(i) holds and suppose $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}$, where $\mathbf{H}$ is symmetric and positive definite. Then $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(1)$ if*

$$\tilde{M}_n^*(\tilde{\boldsymbol{\theta}}_n^*) \geq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \tilde{M}_n^*(\boldsymbol{\theta}) - o_{\mathbb{P}}(1).$$

Proof of Lemma A.6. It suffices to show that every $\delta > 0$ admits a constant $c_\delta^* > 0$ such that

$$\mathbb{P}\left[\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \hat{\boldsymbol{\Theta}}_n^\delta} \tilde{M}_n^*(\boldsymbol{\theta}) > c_\delta^*\right] \to 1, \tag{A.3}$$

where $\hat{\boldsymbol{\Theta}}_n^\delta = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : ||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n|| \leq \delta\}$. The process $\tilde{M}_n^*$ satisfies

$$\tilde{M}_n^*(\boldsymbol{\theta}) = \hat{M}_n^*(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \tilde{\mathbf{H}}_n (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n), \qquad \hat{M}_n^*(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n m_n(\mathbf{z}_{i,n}^*, \boldsymbol{\theta}),$$

where it follows from Pollard (1989, Theorem 4.2) that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\hat{M}_n^*(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta})| = O_{\mathbb{P}}\left(\sqrt{\frac{\mathbb{E}[\bar{m}_n(\mathbf{z})^2]}{n}}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nq_n}}\right) = o_{\mathbb{P}}(1).$$

As a consequence, for any $\delta > 0$,

$$\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \hat{\boldsymbol{\Theta}}_n^\delta} \tilde{M}_n^*(\boldsymbol{\theta}) = \frac{1}{2} \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \hat{\boldsymbol{\Theta}}_n^\delta} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \tilde{\mathbf{H}}_n (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n) + o_{\mathbb{P}}(1),$$

so (A.3) is satisfied with $c_\delta^* = \delta^2 \lambda_{\min}(\mathbf{H})/4 > 0$. $\blacksquare$

Next, because

$$\tilde{M}_n(\boldsymbol{\theta}) = \mathbb{E}_n^*[\tilde{M}_n^*(\boldsymbol{\theta})] = \frac{1}{n} \sum_{i=1}^n \tilde{m}_n(\mathbf{z}_i, \boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)' \tilde{\mathbf{H}}_n (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n),$$

we have the following convergence result about $\tilde{Q}_n$.

Lemma A.7 *Suppose $r_n \to \infty$, $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}$, and suppose $\hat{\boldsymbol{\theta}}_n \to_{\mathbb{P}} \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$. Then $\tilde{Q}_n \to_{\mathbb{P}} \mathcal{Q}$ in the topology of uniform convergence on compacta, where $\mathcal{Q}(\mathbf{s}) = -\mathbf{s}'\mathbf{H}\mathbf{s}/2$;*

that is

$$\sup_{||\mathbf{s}||\le K}\left|\tilde{Q}_n(\mathbf{s})-(-\frac{1}{2}\mathbf{s}'\mathbf{H}\mathbf{s})\right|\to_{\mathbb{P}}0$$

for any $K>0$.

Proof of Lemma A.7. Uniformly in $\mathbf{s}$ with $||\mathbf{s}||\le K$, we have

$$\left|\tilde{Q}_n(\mathbf{s})-(-\frac{1}{2}\mathbf{s}'\mathbf{H}\mathbf{s})\right|\le\frac{1}{2}\left|\mathbf{s}'(\tilde{\mathbf{H}}_n-\mathbf{H})\mathbf{s}\right|+\frac{1}{2}\left|\mathbf{s}'\mathbf{H}\mathbf{s}\right|\mathbb{1}(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1}\notin\boldsymbol{\Theta})\le K^2o_{\mathbb{P}}(1),$$

where the last inequality uses $\tilde{\mathbf{H}}_n\to_{\mathbb{P}}\mathbf{H}$ and $\mathbb{P}(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1}\notin\boldsymbol{\Theta})\to0$. ∎

The next lemma can be used to obtain the rate of convergence of $\tilde{\boldsymbol{\theta}}_n^*$.

Lemma A.8 *Suppose Condition CRA(iii) holds and suppose* $\tilde{\mathbf{H}}_n\to_{\mathbb{P}}\mathbf{H}$, *where* $\mathbf{H}$ *is symmetric and positive definite. Then* $r_n(\tilde{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n)=O_{\mathbb{P}}(1)$ *if* $r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=O_{\mathbb{P}}(1)$, $\tilde{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n=o_{\mathbb{P}}(1)$, *and if*

$$\tilde{M}_n^*(\tilde{\boldsymbol{\theta}}_n^*)\ge\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\tilde{M}_n^*(\boldsymbol{\theta})-o_{\mathbb{P}}(r_n^{-2}).$$

Proof of Lemma A.8. For any $\delta>0$ and any $K\in\mathbb{N}$, $\mathbb{P}[r_n||\tilde{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n||>2^{K+1}]$ is no greater than

$$\mathbb{P}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\tilde{M}_n^*(\boldsymbol{\theta})-\tilde{M}_n^*(\tilde{\boldsymbol{\theta}}_n^*)\ge\delta r_n^{-2}]+\mathbb{P}[||\tilde{\mathbf{H}}_n-\mathbf{H}||>\delta]+\mathbb{P}[||\tilde{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n||>\delta/4]$$
$$+\mathbb{P}[r_n||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||>2^K]$$
$$+\sum_{j\ge K,2^{j+1}\le\delta r_n}\mathbb{P}\left[\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\le2^j,r_n||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||\le2^K,||\tilde{\mathbf{H}}_n-\mathbf{H}||\le\delta}\tilde{M}_n^*(\boldsymbol{\theta})-\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n)\ge-\delta r_n^{-2}\right].$$

By assumption, the probabilities on the first line go to zero for any $\delta>0$ and the probability on the second line can be made arbitrarily small by making K large. As a consequence, it suffices to show that the sum on the last line can be made arbitrarily small (for large n) by making $\delta>0$ small and K large.

To do so, let $\delta>0$ be small enough so that Condition CRA(iii) holds and

$$\frac{1}{2}\inf_{||\bar{\mathbf{H}}-\mathbf{H}||\le\delta}\lambda_{\min}(\bar{\mathbf{H}}+\bar{\mathbf{H}}')>\lambda_{\min}(\mathbf{H}).$$

Then, if $||\tilde{\mathbf{H}}_n-\mathbf{H}||\le\delta$, we have

$$\tilde{M}_n(\hat{\boldsymbol{\theta}}_n)-\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\le2^j}\tilde{M}_n(\boldsymbol{\theta})-\delta r_n^{-2}\ge2^{2j}c_K^*(\delta)r_n^{-2}$$

for any pair $(j,K)'\in\mathbb{N}^2$ with $j\ge K$, where $c_K^*(\delta)=\lambda_{\min}(\mathbf{H})/16-2^{-2K}\delta$.

Choosing K large enough that $c_K^*(\delta)\ge c^*=\lambda_{\min}(\mathbf{H})/32$ and using the fact that

$$\tilde{M}_n^*(\boldsymbol{\theta})-\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n)-\tilde{M}_n(\boldsymbol{\theta})+\tilde{M}_n(\hat{\boldsymbol{\theta}}_n)=\hat{M}_n^*(\boldsymbol{\theta})-\hat{M}_n^*(\hat{\boldsymbol{\theta}}_n)-\hat{M}_n(\boldsymbol{\theta})+\hat{M}_n(\hat{\boldsymbol{\theta}}_n),$$

we therefore have

$$\sum_{j\geq K,2^{j+1}\leq \delta r_n} \mathbb{P}\left[\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\leq 2^j, r_n||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||\leq 2^K, ||\tilde{\mathbf{H}}_n-\mathbf{H}||\leq\delta} \tilde{M}_n^*(\boldsymbol{\theta})-\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n)\geq -\delta r_n^{-2}\right]$$

$$\leq \sum_{j\geq K,2^{j+1}\leq \delta r_n} \mathbb{P}\left[\sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\leq 2^j, r_n||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||\leq 2^K} \{\hat{M}_n^*(\boldsymbol{\theta})-\hat{M}_n^*(\hat{\boldsymbol{\theta}}_n)-\hat{M}_n(\boldsymbol{\theta})+\hat{M}_n(\hat{\boldsymbol{\theta}}_n)\}\geq 2^{2j}c^* r_n^{-2}\right]$$

$$\leq \sum_{j\geq K,2^{j+1}\leq \delta r_n} \mathbb{P}\left[\sup_{r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\leq 2^j, r_n||\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0||\leq 2^K} ||\hat{M}_n^*(\boldsymbol{\theta})-\hat{M}_n^*(\hat{\boldsymbol{\theta}}_n)-\hat{M}_n(\boldsymbol{\theta})+\hat{M}_n(\hat{\boldsymbol{\theta}}_n)||\geq 2^{2j}c^* r_n^{-2}\right]$$

$$\leq \frac{r_n^2}{c^*}\sum_{j\geq K,2^{j+1}\leq \delta r_n} 2^{-2j}\mathbb{E}\left[\sup_{r_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||\leq 2^{j+1}, r_n||\boldsymbol{\theta}'-\boldsymbol{\theta}_0||\leq 2^K} ||\hat{M}_n^*(\boldsymbol{\theta})-\hat{M}_n^*(\boldsymbol{\theta}')-\hat{M}_n(\boldsymbol{\theta})+\hat{M}_n(\boldsymbol{\theta}')||\right],$$

where the last inequality uses the Markov inequality.

Under Condition CRA(iii), $q_n \sup_{0\leq\delta'\leq\delta}\mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$ and Pollard (1989, Theorem 4.2) can be used to show that the sum on the last line is bounded by a constant multiple of

$$r_n^2 \sum_{j\geq K,2^{j+1}\leq\delta r_n} 2^{-2j}\sqrt{\frac{\mathbb{E}[\bar{d}_n^{2^{j+1}/r_n}(\mathbf{z})^2]}{n}} \leq \sqrt{2q_n \sup_{0\leq\delta'\leq\delta}\mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta']}\sum_{j\geq K}2^{-3j/2},$$

which can be made arbitrarily small by making K large. ∎

Finally, the next two lemmas can be combined to show that $\tilde{G}_n^* \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0$ in the topology of uniform convergence on compacta.

Lemma A.9 *Suppose Conditions CRA(iii)-(iv) hold, $r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=O_{\mathbb{P}}(1)$, and that, for every $K>0$, $\sup_{||\mathbf{s}||\leq K}|\hat{G}_n(\mathbf{s})+Q_n(\mathbf{s})| = o_{\mathbb{P}}(\sqrt{n})$. Then $\tilde{G}_n^*$ converges to $\mathcal{G}_0$ in the sense of conditional weak convergence in probability of finite-dimensional projections.*

Proof of Lemma A.9. Because $\tilde{G}_n^*(\mathbf{s}) = n^{-1/2}\sum_{i=1}^n \hat{\psi}_n(\mathbf{z}_{i,n}^*;\mathbf{s})$, where

$$\hat{\psi}_n(\mathbf{z};\mathbf{s}) = \sqrt{r_n q_n}[m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n)-\hat{M}_n(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})+\hat{M}_n(\hat{\boldsymbol{\theta}}_n)]\mathbb{1}(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1}\in\boldsymbol{\Theta}),$$

the result follows from the Cramér-Wold device if

$$\mathbb{E}_n^*[\hat{\psi}_n(\mathbf{z}^*;\mathbf{s})\hat{\psi}_n(\mathbf{z}^*;\mathbf{t})] = \frac{1}{n}\sum_{i=1}^n \hat{\psi}_n(\mathbf{z}_i;\mathbf{s})\hat{\psi}_n(\mathbf{z}_i;\mathbf{t}) \to_{\mathbb{P}} \mathcal{C}_0(\mathbf{s},\mathbf{t}) \qquad \forall \mathbf{s},\mathbf{t}\in\mathbb{R}^d,$$

and if the following Lyapunov condition is satisfied:

$$\frac{1}{n}\mathbb{E}_n^*[\hat{\psi}_n(\mathbf{z}^*;\mathbf{s})^4] = \frac{1}{n^2}\sum_{i=1}^n \hat{\psi}_n(\mathbf{z}_i;\mathbf{s})^4 \to_{\mathbb{P}} 0 \qquad \forall \mathbf{s}\in\mathbb{R}^d.$$

Let $\mathbf{s},\mathbf{t}\in\mathbb{R}^d$ be given and suppose without loss of generality that $\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1}, \hat{\boldsymbol{\theta}}_n+\mathbf{t}r_n^{-1}\in\boldsymbol{\Theta}$. Because $r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=O_{\mathbb{P}}(1)$, we have

$$\begin{aligned}\hat{Q}_n(\mathbf{s}) &= r_n^2[\hat{M}_n(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-\hat{M}_n(\hat{\boldsymbol{\theta}}_n)]\mathbb{1}(\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1}\in\boldsymbol{\Theta})\\ &= \{\hat{G}_n[r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+\mathbf{s}]+Q_n[r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+\mathbf{s}]\}-\{\hat{G}_n[r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)]+Q_n[r_n(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)]\}\\ &= o_{\mathbb{P}}(\sqrt{n})\end{aligned}$$

and, using $\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = o(q_n^{-3} r_n)$ (for $\delta_n = O(r_n^{-1})$) and Pollard (1989, Theorem 4.2),

$$\begin{aligned} & r_n q_n \mathbb{E}_n^*[\{m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n)\}\{m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{t} r_n^{-1}) - m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n)\}] - \hat{\mathcal{C}}_n(\mathbf{s},\mathbf{t}) \\ = & \frac{r_n q_n}{n} \sum_{i=1}^n \{m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n)\}\{m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n + \mathbf{t} r_n^{-1}) - m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n)\} - \hat{\mathcal{C}}_n(\mathbf{s},\mathbf{t}) \\ = & o_{\mathbb{P}}\left(r_n q_n \sqrt{\frac{r_n}{n q_n^3}} \right) = o_{\mathbb{P}}(1), \end{aligned}$$

where

$$\begin{aligned} \hat{\mathcal{C}}_n(\mathbf{s},\mathbf{t}) &= r_n q_n \, \mathbb{E}[\{m_n(\mathbf{z}, \boldsymbol{\theta} + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}, \boldsymbol{\theta})\}\{m_n(\mathbf{z}, \boldsymbol{\theta} + \mathbf{t} r_n^{-1}) - m_n(\mathbf{z}, \boldsymbol{\theta})\}]\big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n} \\ &= \mathcal{C}_0(\mathbf{s},\mathbf{t}) + o_{\mathbb{P}}(1). \end{aligned}$$

Using these facts and the representation

$$\hat{\psi}_n(\mathbf{z}; \mathbf{s}) = \sqrt{r_n q_n}[m_n(\mathbf{z}, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}, \hat{\boldsymbol{\theta}}_n)] - \frac{1}{\sqrt{n}} \hat{Q}_n(\mathbf{s}),$$

we have

$$\begin{aligned} & \mathbb{E}_n^*[\hat{\psi}_n(\mathbf{z}^*; \mathbf{s}) \hat{\psi}_n(\mathbf{z}^*; \mathbf{t})] \\ = & r_n q_n \mathbb{E}_n^*[\{m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n)\}\{m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{t} r_n^{-1}) - m_n(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n)\}] - \frac{1}{n} \hat{Q}_n(\mathbf{s}) \hat{Q}_n(\mathbf{t}) \\ = & \mathcal{C}_0(\mathbf{s},\mathbf{t}) + o_{\mathbb{P}}(1) \end{aligned}$$

and, using $\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = o(q_n^{-3} r_n)$ (for $\delta_n = O(r_n^{-1})$),

$$\begin{aligned} \frac{1}{16n} \mathbb{E}_n^*[\hat{\psi}_n(\mathbf{z}^*; \mathbf{s})^4] &= \frac{1}{16n^2} \sum_{i=1}^n \hat{\psi}_n(\mathbf{z}_i; \mathbf{s})^4 \leq \frac{r_n^2 q_n^2}{n^2} \sum_{i=1}^n |m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n)|^4 + \frac{1}{n^3} \hat{Q}_n(\mathbf{s})^4 \\ &= o_{\mathbb{P}}\left(\frac{r_n^3}{n q_n} + \frac{1}{n} \right) = o_{\mathbb{P}}(1). \quad \blacksquare \end{aligned}$$

Lemma A.10 *Suppose Conditions CRA(iii) and CRA(v) hold and suppose* $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$. *Then* $\{\tilde{G}_n^*(\mathbf{s}) : ||\mathbf{s}|| \leq K\}$ *is stochastically equicontinuous for every* $K > 0$; *that is,*

$$\sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||, ||\mathbf{t}|| \leq K}} |\tilde{G}_n^*(\mathbf{s}) - \tilde{G}_n^*(\mathbf{t})| \to_{\mathbb{P}} 0$$

for any $K > 0$ *and for any* $\Delta_n > 0$ *with* $\Delta_n = o(1)$.

Proof of Lemma A.10. Let $K > 0$ be given. Proceeding as in the proof of Kim and Pollard (1990, Lemma 4.6) and using $q_n \delta_n^{-1} \mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^2] = O(1)$ (for $\delta_n = O(r_n^{-1})$) along with the fact that $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$, it suffices to show that, for every finite $k > 0$,

$$\begin{aligned} & r_n \mathbb{1}\left(r_n ||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| \leq k \right) \sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||, ||\mathbf{t}|| \leq K}} \frac{q_n}{n} \sum_{i=1}^n \hat{d}_n(\mathbf{z}_{i,n}^*; \mathbf{s}, \mathbf{t})^2 \\ \leq & r_n \sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||, ||\mathbf{t}|| \leq K+k}} \frac{q_n}{n} \sum_{i=1}^n d_n(\mathbf{z}_{i,n}^*; \mathbf{s}, \mathbf{t})^2 \to_{\mathbb{P}} 0, \end{aligned}$$

where

$$\hat{d}_n(\mathbf{z};\mathbf{s},\mathbf{t}) = \frac{1}{2}|m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1}) - m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n + \mathbf{t}r_n^{-1})| = d_n(\mathbf{z}; r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{s}, r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{t}).$$

Let $k > 0$ be given. For any $C > 0$ and any $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$ with $||\mathbf{s}||, ||\mathbf{t}|| \leq K + k$,

$$\begin{aligned}
\frac{q_n}{n}\sum_{i=1}^n d_n(\mathbf{z}_{i,n}^*;\mathbf{s},\mathbf{t})^2 \quad \leq \quad & \frac{q_n}{n}\sum_{i=1}^n \bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z}_{i,n}^*)^2 \mathbb{1}(q_n \bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z}_{i,n}^*) > C) \\
& + C\mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})] \\
& + C\frac{1}{n}\sum_{i=1}^n \{d_n(\mathbf{z}_{i,n};\mathbf{s},\mathbf{t}) - \mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]\} \\
& + C\frac{1}{n}\sum_{i=1}^n \{d_n(\mathbf{z}_{i,n}^*;\mathbf{s},\mathbf{t}) - \mathbb{E}_n^*[d_n(\mathbf{z}^*;\mathbf{s},\mathbf{t})]\},
\end{aligned}$$

and therefore

$$\begin{aligned}
& r_n \mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||,||\mathbf{t}|| \leq K+k}} \frac{q_n}{n}\sum_{i=1}^n d_n(\mathbf{z}_{i,n}^*;\mathbf{s},\mathbf{t})^2\right] \\
\leq \quad & r_n \mathbb{E}\left[\frac{q_n}{n}\sum_{i=1}^n \bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z}_{i,n}^*)^2 \mathbb{1}(q_n \bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z}_{i,n}^*) > C)\right] \\
& + Cr_n \sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||,||\mathbf{t}|| \leq K+k}} \mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})] \\
& + Cr_n \mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||,||\mathbf{t}|| \leq K+k}} \left|\frac{1}{n}\sum_{i=1}^n \{d_n(\mathbf{z}_{i,n};\mathbf{s},\mathbf{t}) - \mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]\}\right|\right] \\
& + Cr_n \mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}|| \leq \Delta_n \\ ||\mathbf{s}||,||\mathbf{t}|| \leq K+k}} \left|\frac{1}{n}\sum_{i=1}^n \{d_n(\mathbf{z}_{i,n}^*;\mathbf{s},\mathbf{t}) - \mathbb{E}^*[d_n(\mathbf{z}^*;\mathbf{s},\mathbf{t})]\}\right|\right].
\end{aligned}$$

For large n, the first term on the majorant side can be made arbitrarily small by making C large. Also, for any fixed C, the second term tends to zero because $\Delta_n \to 0$. Finally, Pollard (1989, Theorem 4.2) can be used to show that for fixed C and for large n, each of the last two terms is bounded by a constant multiple of

$$r_n\sqrt{\frac{\mathbb{E}[\bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z})^2]}{n}} = \frac{\sqrt{K+k}}{r_n}\sqrt{q_n\mathbb{E}[\bar{d}_n^{(K+k)r_n^{-1}}(\mathbf{z})^2/\{(K+k)r_n^{-1}\}]} = O\left(\frac{1}{r_n}\right) = o(1). \quad \blacksquare$$

A.1.2 Proof of Lemma 1

Without loss of generality, suppose $r_n||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| \leq K$ for some fixed constant K. Defining

$$\check{H}_{n,kl}^{\mathtt{ND}} = -\frac{1}{4\epsilon_n^2}[\hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l) - \hat{M}_n(\boldsymbol{\theta}_0-\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l) - \hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l) + \hat{M}_n(\boldsymbol{\theta}_0-\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l)]$$

and

$$\bar{H}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}) = -\frac{1}{4\epsilon_n^2}[M_n(\boldsymbol{\theta}+\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l)-M_n(\boldsymbol{\theta}-\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l)-M_n(\boldsymbol{\theta}+\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l)+M_n(\boldsymbol{\theta}-\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l)],$$

we obtain the decomposition

$$\tilde{H}^{\texttt{ND}}_{n,kl} = \check{H}^{\texttt{ND}}_{n,kl} + R^{\texttt{ND}}_{n,kl} + S^{\texttt{ND}}_{n,kl},$$

where

$$R^{\texttt{ND}}_{n,kl} = \tilde{H}^{\texttt{ND}}_{n,kl} - \check{H}^{\texttt{ND}}_{n,kl} - \bar{H}^{\texttt{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) + \bar{H}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0), \qquad S^{\texttt{ND}}_{n,kl} = \bar{H}^{\texttt{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) - \bar{H}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0).$$

The proof will be completed by showing that $\check{H}^{\texttt{ND}}_{n,kl} \to_{\mathbb{P}} H_{0,kl}$, $R^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}(1)$, and $S^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}(1)$.

First, using (A.2) and the fact that $\dot{C}_n = o(r_n^{-1})$ and $\ddot{C}_n = o(1)$ under Condition CRA(ii), we have

$$M_n(\boldsymbol{\theta}_0 + \epsilon_n\mathbf{e}_k + \epsilon_n\mathbf{e}_l) - M_n(\boldsymbol{\theta}_0) = -\epsilon_n^2\frac{1}{2}(\mathbf{e}_k + \mathbf{e}_l)'\mathbf{H}_n(\mathbf{e}_k + \mathbf{e}_l) + o\left(\frac{\epsilon_n}{r_n} + \epsilon_n^2\right),$$

implying in particular that

$$\bar{H}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0) = H_{n,kl} + o\left(\frac{1}{r_n\epsilon_n} + 1\right),$$

where, using $\mathbf{H}_n \to \mathbf{H}_0$,

$$H_{n,kl} = \mathbf{e}_k'\mathbf{H}_n\mathbf{e}_l \to \mathbf{e}_k'\mathbf{H}_0\mathbf{e}_l = H_{0,kl}.$$

Moreover, $\check{H}^{\texttt{ND}}_{n,kl} - \bar{H}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0)$ is $o_{\mathbb{P}}(1)$ because it has mean zero and its variance is bounded by a constant multiple of

$$\frac{\mathbb{E}[\bar{d}_n^{2\epsilon_n}(\mathbf{z})^2]}{n\epsilon_n^4} = O\left(\frac{1}{nq_n\epsilon_n^3}\right) = O\left(\frac{1}{r_n^3\epsilon_n^3}\right) = o(1).$$

As a consequence, $\check{H}^{\texttt{ND}}_{n,kl} \to_{\mathbb{P}} H_{0,kl}$.

Next, to show that $R^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}(1)$ it suffices to show that

$$\frac{1}{\epsilon_n^2}\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq Kr_n^{-1}+2\epsilon_n}|\hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}_0)| = o_{\mathbb{P}}(1).$$

The displayed result holds because it follows from Pollard (1989, Theorem 4.2) that

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\epsilon_n^2}\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\leq Kr_n^{-1}+2\epsilon_n}|\hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}) + M_n(\boldsymbol{\theta}_0)|\right] &= O\left(\sqrt{\frac{\mathbb{E}[\bar{d}_n^{Cr_n^{-1}+2\epsilon_n}(\mathbf{z})^2]}{n\epsilon_n^4}}\right) \\ &= O\left(\frac{1}{\sqrt{r_n^3\epsilon_n^3}}\right) = o(1).\end{aligned}$$

Finally, making repeated use of (A.2) and the fact that $r_n||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| \leq K$, we have

$$S^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}\left(\frac{1}{r_n^2\epsilon_n^2} + 1\right) = o_{\mathbb{P}}(1).$$

A.1.3 Proof of Lemma 2

Letting $\check{H}^{\texttt{ND}}_{n,kl}$, $R^{\texttt{ND}}_{n,kl}$, and $S^{\texttt{ND}}_{n,kl}$ be defined as in the proof of Lemma 1, we have $R^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}(1/\sqrt{r_n^3\epsilon_n^3})$ because Pollard (1989, Theorem 4.2) can be used to show that for any $K > 0$ and for any $\Delta_n > 0$ with $\Delta_n = o(1)$,

$$\begin{aligned}
&\frac{1}{\epsilon_n^2} \sup_{\substack{\|\mathbf{s}-\mathbf{t}\|\leq\Delta_n \\ \|\mathbf{s}\|,\|\mathbf{t}\|\leq K}} \left|\hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{s}) - \hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{t}) - M_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{s}) + M_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{t})\right| \\
&= \frac{1}{\epsilon_n^2} o_{\mathbb{P}}\left(\frac{\sqrt{r_n\epsilon_n}}{r_n^2}\right) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{r_n^3\epsilon_n^3}}\right).
\end{aligned}$$

Also, Taylor's theorem can be used to show that

$$S^{\texttt{ND}}_{n,kl} = -\{\frac{\partial}{\partial\boldsymbol{\theta}}\ddot{M}_{n,kl}(\boldsymbol{\theta}_0)\}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{\mathbb{P}}(\epsilon_n^2).$$

As a consequence, $\tilde{H}^{\texttt{ND}}_{n,kl} - \check{H}^{\texttt{ND}}_{n,kl} = o_{\mathbb{P}}(\epsilon_n^2 + 1/\sqrt{r_n^3\epsilon_n^3}) + O_{\mathbb{P}}(1/r_n)$, where the $O_{\mathbb{P}}(1/r_n)$ term

$$-\{\frac{\partial}{\partial\boldsymbol{\theta}}\ddot{M}_{n,kl}(\boldsymbol{\theta}_0)\}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

does not depend on ϵ_n.

Next, we approximate the moments of $\check{H}^{\texttt{ND}}_{n,kl}$. First, using Taylor's theorem, it can be shown that

$$\mathbb{E}[\check{H}^{\texttt{ND}}_{n,kl}] - H_{n,kl} = -\epsilon_n^2\mathsf{B}_{n,kl} + o(\epsilon_n^2),$$

where

$$\mathsf{B}_{n,kl} = -\frac{1}{6}\left[\frac{\partial^2}{\partial\theta_k^2}\ddot{M}_{n,kl}(\boldsymbol{\theta}_0) + \frac{\partial^2}{\partial\theta_l^2}\ddot{M}_{n,kl}(\boldsymbol{\theta}_0)\right] \to -\frac{1}{6}\left[\frac{\partial^2}{\partial\theta_k^2}\ddot{M}_{0,kl}(\boldsymbol{\theta}_0) + \frac{\partial^2}{\partial\theta_l^2}\ddot{M}_{0,kl}(\boldsymbol{\theta}_0)\right] = \mathsf{B}_{kl}.$$

Finally, to obtain an expression for the variance of $\check{H}^{\texttt{ND}}_{n,kl}$, let $m^{\Delta}_{n,kl}(\mathbf{z})$ denote

$$m_n(\mathbf{z},\boldsymbol{\theta}_0+\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l) - m_n(\mathbf{z},\boldsymbol{\theta}_0+\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l) - m_n(\mathbf{z},\boldsymbol{\theta}_0-\epsilon_n\mathbf{e}_k+\epsilon_n\mathbf{e}_l) + m_n(\mathbf{z},\boldsymbol{\theta}_0-\epsilon_n\mathbf{e}_k-\epsilon_n\mathbf{e}_l).$$

Because

$$\check{H}^{\texttt{ND}}_{n,kl} = -\frac{1}{4n\epsilon_n^2}\sum_{i=1}^n m^{\Delta}_{n,kl}(\mathbf{z}_i),$$

we have

$$\mathbb{V}[\check{H}^{\texttt{ND}}_{n,kl}] = \frac{1}{16n\epsilon_n^4}\mathbb{V}[m^{\Delta}_{n,kl}(\mathbf{z})] = \frac{1}{16n\epsilon_n^4}\mathbb{E}[m^{\Delta}_{n,kl}(\mathbf{z})^2] + O\left(\frac{1}{n}\right).$$

Also, by condition CRA(iv),

$$\frac{q_n}{\epsilon_n}\mathbb{E}[\{m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}\epsilon_n) - m_n(\mathbf{z},\boldsymbol{\theta}_0)\}\{m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{t}\epsilon_n) - m_n(\mathbf{z},\boldsymbol{\theta}_0)\}] \to \mathcal{C}_0(\mathbf{s},\mathbf{t}).$$

Therefore,

$$\mathbb{V}[\check{H}^{\texttt{ND}}_{n,kl}] = \frac{1}{r_n^3\epsilon_n^3}[\mathsf{V}_{n,kl} + o(1)] + O\left(\frac{1}{n}\right) = \frac{1}{r_n^3\epsilon_n^3}\mathsf{V}_{kl} + o\left(\frac{1}{r_n^3\epsilon_n^3}\right),$$

where, using $\mathcal{C}_0(\mathbf{s},-\mathbf{s})=0$ and $\mathcal{C}_0(\mathbf{s},\mathbf{t})=\mathcal{C}_0(-\mathbf{s},-\mathbf{t})$,

$$\begin{aligned}\mathsf{V}_{n,kl} &= \frac{q_n}{16\epsilon_n}\mathbb{E}[m^{\Delta}_{n,kl}(\mathbf{z})^2] \\ &\to \frac{1}{8}[\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,\mathbf{e}_k+\mathbf{e}_l)+\mathcal{C}_0(\mathbf{e}_k-\mathbf{e}_l,\mathbf{e}_k-\mathbf{e}_l)-2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,\mathbf{e}_k-\mathbf{e}_l)-2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,-\mathbf{e}_k+\mathbf{e}_l)] \\ &= \mathsf{V}_{kl}.\end{aligned}$$

A.1.4 The Benchmark Case

The remainder of the supplemental appendix verifies Condition CRA for the four examples in the paper. In three of those examples (namely, maximum score, panel maximum score, and empirical risk minimization), the function m_n does not depend on n. To state a simplified version of Condition CRA applicable in such cases, let the function m_n be denoted by m_0 and for any $\delta>0$, define

$$\bar{m}_0(\mathbf{z})=\sup_{m\in\mathcal{M}_0}|m(\mathbf{z})|, \qquad \mathcal{M}_0=\{m_0(\cdot,\boldsymbol{\theta}):\boldsymbol{\theta}\in\boldsymbol{\Theta}\},$$

and

$$\bar{d}^{\delta}_0(\mathbf{z})=\sup_{d\in\mathcal{D}^{\delta}_0}|d(\mathbf{z})|, \qquad \mathcal{D}^{\delta}_0=\{m_0(\cdot,\boldsymbol{\theta})-m_0(\cdot,\boldsymbol{\theta}_0):\boldsymbol{\theta}\in\boldsymbol{\Theta}^{\delta}_0\}.$$

Condition CRA$_0$ (Cube Root Asymptotics, benchmark case) The following are satisfied:

(i) $\mathcal{M}_0$ is manageable for the envelope $\bar{m}_0$ and $\mathbb{E}[\bar{m}_0(\mathbf{z})^2]<\infty$.
Also, for every $\delta>0$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}^{\delta}_0}M_0(\boldsymbol{\theta})<M_0(\boldsymbol{\theta}_0)$.

(ii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$ and, for some $\delta>0$, M_0 is twice continuously differentiable on $\boldsymbol{\Theta}^{\delta}_0$. Also, $\mathbf{H}_0=-\partial^2M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(iii) For some $\delta>0$, $\{\mathcal{D}^{\delta'}_0:0<\delta'\leq\delta\}$ is uniformly manageable for the envelopes $\bar{d}^{\delta'}_0$ and $\sup_{0<\delta'\leq\delta}\mathbb{E}[\bar{d}^{\delta'}_0(\mathbf{z})^2/\delta']<\infty$.

(iv) For every $\delta_n>0$ with $\delta_n=O(n^{-1/3})$, $n^{-1/3}\mathbb{E}[\bar{d}^{\delta_n}_0(\mathbf{z})^4]=o(1)$ and, for all $\mathbf{s},\mathbf{t}\in\mathbb{R}^d$ and for some $\mathcal{C}_0$ with $\mathcal{C}_0(\mathbf{s},\mathbf{s})+\mathcal{C}_0(\mathbf{t},\mathbf{t})-2\mathcal{C}_0(\mathbf{s},\mathbf{t})>0$ for $\mathbf{s}\neq\mathbf{t}$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}^{\delta_n}_0}\left|\frac{1}{\delta_n}\mathbb{E}[\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s})-m_0(\mathbf{z},\boldsymbol{\theta})\}\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t})-m_0(\mathbf{z},\boldsymbol{\theta})\}]-\mathcal{C}_0(\mathbf{s},\mathbf{t})\right|=o(1).$$

(v) For every $\delta_n>0$ with $\delta_n=O(n^{-1/3})$,

$$\lim_{C\to\infty}\limsup_{n\to\infty}\sup_{0<\delta\leq\delta_n}\mathbb{E}[\mathbb{1}(\bar{d}^{\delta}_0(\mathbf{z})>C)\bar{d}^{\delta}_0(\mathbf{z})^2/\delta]=0$$

and $\sup_{\boldsymbol{\theta},\boldsymbol{\theta}'\in\boldsymbol{\Theta}^{\delta_n}_0}\mathbb{E}[|m_0(\mathbf{z},\boldsymbol{\theta})-m_0(\mathbf{z},\boldsymbol{\theta}')|]/||\boldsymbol{\theta}-\boldsymbol{\theta}'||=O(1)$.

Lemma A.11 *If Condition CRA$_0$ is satisfied, then Condition CRA is satisfied with $q_n=1$.*

A.2 Example: Maximum Score

To state sufficient conditions for Condition CRA$_0$ in this example, let $F_{a|\mathbf{b}}$ denote the conditional distribution function of a given $\mathbf{b}$.

Condition MS For some $\delta > 0$, $S_F \geq 1$, and $S_M \geq 2$, the following are satisfied:

(i) $0 < \mathbb{P}(y = 1|\mathbf{x}) < 1$ almost surely and $F_{u|x_1,\mathbf{x}_2}(u|x_1, \mathbf{x}_2)$ is S_F times continuously differentiable in u and x_1 with bounded derivatives.

(ii) The support of $\mathbf{x}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+1}$, $\mathbb{E}[\|\mathbf{x}_2\|^2] < \infty$, and conditional on $\mathbf{x}_2$, x_1 has everywhere positive Lebesgue density. Also, $F_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)$ is S_F times continuously differentiable in x_1 with bounded derivatives.

(iii) $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

(iv) $M^{\texttt{MS}}(\boldsymbol{\theta}) = \mathbb{E}[m^{\texttt{MS}}(\mathbf{z}, \boldsymbol{\theta})]$ is S_M times continuously differentiable in $\boldsymbol{\theta}$ on $\boldsymbol{\Theta}_0^\delta$ and

$$\mathbf{H}^{\texttt{MS}} = 2\mathbb{E}[f_{u|x_1,\mathbf{x}_2}(0| - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2']$$

is positive definite.

Corollary MS *Suppose Condition MS is satisfied. Then Condition CRA is satisfied with* $q_n = 1$, $\mathbf{H}_0 = \mathbf{H}^{\texttt{MS}}$, *and* $\mathcal{C}_0 = \mathcal{C}^{\texttt{MS}}$, *where*

$$\mathcal{C}^{\texttt{MS}}(\mathbf{s}, \mathbf{t}) = \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} \mathbb{1}(\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t}))].$$

Alternative representations of $\mathbf{H}^{\texttt{MS}}$ and $\mathcal{C}^{\texttt{MS}}$ are available. In particular, defining

$$\begin{aligned} \eta^{\texttt{MS}}(\mathbf{x}_2) &= \left\{ \frac{\partial}{\partial x_1} \mathbb{E}(2y - 1|x_1, \mathbf{x}_2) \right\} f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2) \Big|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0} \\ &= 2 f_{u|x_1,\mathbf{x}_2}(0| - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) \end{aligned}$$

and

$$\begin{aligned} \psi^{\texttt{MS}}(\mathbf{x}_2) &= \mathbb{E}[(2y-1)^2|x_1, \mathbf{x}_2] f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)\big|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0} \\ &= f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2), \end{aligned}$$

we have

$$\mathbf{H}^{\texttt{MS}} = \mathbb{E}[\eta^{\texttt{MS}}(\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2']$$

and

$$\mathcal{C}^{\texttt{MS}}(\mathbf{s}, \mathbf{t}) = \mathbb{E}[\psi^{\texttt{MS}}(\mathbf{x}_2) \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} \mathbb{1}(\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t}))].$$

Similar representations will be obtained for the other two maximum score examples.

As an estimator of $\mathbf{H}^{\texttt{MS}}$, the generic numerical derivative estimator can be used directly. Another option is to employ a "plug-in" estimator, where the conditional densities are replaced by nonparametric estimators thereof. As a third alternative, consider the example-specific construction $\tilde{\mathbf{H}}_n^{\texttt{MS}}$ discussed in the paper. To obtain results for that estimator, we impose some standard conditions on the (derivative of the) kernel function.

Condition K The following are satisfied:

(i) $\int_{\mathbb{R}} \dot{K}(u)^2 du + \int_{\mathbb{R}} (1 + |u|^3)|\dot{K}(u)| du < \infty$.

(ii) $\int_{\mathbb{R}} \dot{K}(u) du = 0$, $\int_{\mathbb{R}} u\dot{K}(u) du = -1$, and $\int_{\mathbb{R}} u^2 \dot{K}(u) du = 0$.

(iii) $\int_{\mathbb{R}} \bar{K}(u)^2 du < \infty$, where $\bar{K}(u) = \sup_{v \neq u} |\dot{K}(v) - \dot{K}(u)|/|v - u|$.

Under Condition K, $\tilde{\mathbf{H}}_n^{\mathtt{MS}}$ admits counterparts of Lemmas 1 and 2 in the paper. To state these, we let $\tilde{H}_{n,kl}^{\mathtt{MS}}$ and $H_{kl}^{\mathtt{MS}}$ denote element (k,l) of $\tilde{\mathbf{H}}_n^{\mathtt{MS}}$ and $\mathbf{H}^{\mathtt{MS}}$, respectively, and define

$$\mathsf{B}_{kl} = \mathbb{E}[\{F_0^{(1,3)}(\mathbf{x}_2) + F_0^{(2,2)}(\mathbf{x}_2) + F_0^{(3,1)}(\mathbf{x}_2)/3\}x_{2,k}x_{2,l}] \int_{\mathbb{R}} u^3 \dot{K}(u)du$$

and

$$\mathsf{V}_{kl} = 2\mathbb{E}[F_0^{(0,1)}(\mathbf{x}_2)x_{2,k}^2 x_{2,l}^2] \int_{\mathbb{R}} \dot{K}(u)^2 du,$$

where $x_{2,k} = \mathbf{e}_k'\mathbf{x}_2$ and

$$F_0^{(i,j)}(\mathbf{x}_2) = \frac{\partial^i}{\partial u^i} F_{u|x_1,\mathbf{x}_2}(-u|x_1+u,\mathbf{x}_2) \frac{\partial^j}{\partial x_1^j} F_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)\bigg|_{u=0,x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0}.$$

Lemma MS *Suppose Conditions MS and K hold.*
(i) If $h_n \to 0$, $nh_n^3 \to \infty$, *and if* $\mathbb{E}[\|\mathbf{x}_2\|^6] < \infty$, *then* $\tilde{\mathbf{H}}_n^{\mathtt{MS}} \to_{\mathbb{P}} \mathbf{H}^{\mathtt{MS}}$.
(ii) If also $S_F \geq 3$ *and* $S_M \geq 4$, *then* $\tilde{H}_{n,kl}^{\mathtt{MS}}$ *admits an approximation* $\check{H}_{n,kl}^{\mathtt{MS}}$ *satisfying*

$$\tilde{H}_{n,kl}^{\mathtt{MS}} = \check{H}_{n,kl}^{\mathtt{MS}} + o_{\mathbb{P}}\left(h_n^2 + \frac{1}{\sqrt{nh_n^3}}\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt[3]{n}}\right)$$

where the $O_{\mathbb{P}}(1/\sqrt[3]{n})$ *term does not depend on* h_n, *and where*

$$\mathbb{E}[(\check{H}_{n,kl}^{\mathtt{MS}} - H_{kl}^{\mathtt{MS}})^2] = h_n^4\mathsf{B}_{kl}^2 + \frac{1}{nh_n^3}\mathsf{V}_{kl} + o\left(h_n^4 + \frac{1}{nh_n^3}\right).$$

A.2.1 Proof of Corollary MS

By Lemma A.11, it suffices to verify that Condition CRA_0 is satisfied.

Condition CRA_0(i). The manageability assumption can be verified using the same argument as in Kim and Pollard (1990). Note that the function $|m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta})|$ is bounded by unity in this example, and thus finite second moment condition holds. It is easy to show that $\boldsymbol{\theta}_0$ uniquely maximizes $M_0(\boldsymbol{\theta})$ over the parameter set. Well-separatedness follows from unique maximum, compactness of the parameter space, and continuity of the function $M_0(\boldsymbol{\theta})$.

Condition CRA_0(ii). Conditions MS(iii)-(iv) imply this condition with $\mathbf{H}_0 = \mathbf{H}^{\mathtt{MS}}$.

Condition CRA_0(iii). Uniform manageability can be verified using the same argument as in Kim and Pollard (1990). Note $d_0^\delta(\mathbf{z}) = \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta} |\mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0) - \mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0 \geq 0)|$. The condition $\sup_{0<\delta'\leq\delta} \mathbb{E}[\bar{d}_0^{\delta'}(\mathbf{z})]/\delta' < \infty$ is verified in Abrevaya and Huang (2005).

Condition CRA_0(iv). Since $d_0^\delta(\mathbf{z})^4 = d_0^\delta(\mathbf{z})$, $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^4] = O(\delta_n)$, which implies the first condition. Also,

$$\mathcal{C}^{\mathtt{MS}}(\mathbf{s},\mathbf{t}) = \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\min\{|\mathbf{x}_2'\mathbf{s}|,|\mathbf{x}_2'\mathbf{t}|\}\mathbb{1}(\text{sgn}(\mathbf{x}_2'\mathbf{s}) = \text{sgn}(\mathbf{x}_2'\mathbf{t}))]$$

satisfies $\mathcal{C}^{\mathtt{MS}}(\mathbf{s},\mathbf{s}) + \mathcal{C}^{\mathtt{MS}}(\mathbf{t},\mathbf{t}) - 2\mathcal{C}^{\mathtt{MS}}(\mathbf{s},\mathbf{t}) > 0$ for $\mathbf{s} \neq \mathbf{t}$. Finally, $\mathcal{C}^{\mathtt{MS}}$ admits the representation

$$\mathcal{C}^{\mathtt{MS}}(\mathbf{s},\mathbf{t}) = \frac{1}{2}[\mathcal{B}^{\mathtt{MS}}(\mathbf{s}) + \mathcal{B}^{\mathtt{MS}}(\mathbf{t}) - \mathcal{B}^{\mathtt{MS}}(\mathbf{s}-\mathbf{t})], \qquad \mathcal{B}^{\mathtt{MS}}(\mathbf{s}) = \mathbb{E}\left[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)|\mathbf{x}_2'\mathbf{s}|\right].$$

Using this representation and the fact that $2xy = x^2 + y^2 - (x-y)^2$, the displayed part of Condition CRA$_0$(iv) can be verified with $\mathcal{C}_0 = \mathcal{C}^{\mathtt{MS}}$ by showing that for $\delta_n = O(n^{-1/3})$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}} \left| \frac{1}{\delta_n}\mathbb{E}|m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s}) - m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t})|^2 - \mathcal{B}^{\mathtt{MS}}(\mathbf{s}-\mathbf{t}) \right| = o(1).$$

Defining $\boldsymbol{\theta}_{\mathbf{s},n} = \boldsymbol{\theta} + \delta_n\mathbf{s}$ and $\boldsymbol{\theta}_{\mathbf{t},n} = \boldsymbol{\theta} + \delta_n\mathbf{t}$, we have, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{\delta_n}$,

$$\begin{aligned}
&\frac{1}{\delta_n}\mathbb{E}|m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}_{\mathbf{s},n}) - m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}_{\mathbf{t},n})|^2 \\
&= \frac{1}{\delta_n}\mathbb{E}\left[\mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{s},n} \geq 0 > x_1 + \mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{t},n})\right] + \frac{1}{\delta_n}\mathbb{E}\left[\mathbb{1}\left(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{t},n} \geq 0 > x_1 + \mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{s},n}\right)\right] \\
&= \frac{1}{\delta_n}\mathbb{E}\left[\int_{-\mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{s},n}}^{-\mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{t},n}} f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)dx_1\mathbb{1}\left(\mathbf{x}_2'\mathbf{t} < \mathbf{x}_2'\mathbf{s}\right)\right] + \frac{1}{\delta_n}\mathbb{E}\left[\int_{-\mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{t},n}}^{-\mathbf{x}_2'\boldsymbol{\theta}_{\mathbf{s},n}} f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)dx_1\mathbb{1}\left(\mathbf{x}_2'\mathbf{s} < \mathbf{x}_2'\mathbf{t}\right)\right] \\
&= \mathbb{E}\left[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2)\mathbf{x}_2'(\mathbf{s}-\mathbf{t})\mathbb{1}\left(\mathbf{x}_2'\mathbf{t} < \mathbf{x}_2'\mathbf{s}\right)\right] + \mathbb{E}\left[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2)\mathbf{x}_2'(\mathbf{t}-\mathbf{s})\mathbb{1}\left(\mathbf{x}_2'\mathbf{s} < \mathbf{x}_2'\mathbf{t}\right)\right] + o(1) \\
&= \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)|\mathbf{x}_2'\mathbf{s} - \mathbf{x}_2'\mathbf{t}|] + o(1),
\end{aligned}$$

from which the desired result follows.

Condition CRA$_0$(v). The first part easily follows from $\bar{d}_0^{\delta}(\mathbf{z}) \leq 1$, while the second part follows from the verification of Condition CRA$_0$(iv).

A.2.2 Proof of Lemma MS

A.2.2.1 Part (i) [Consistency]

Defining

$$\check{\mathbf{H}}_n^{\mathtt{MS}} = -\frac{1}{n}\sum_{i=1}^n (2y_i - 1)\dot{K}_n(x_{1i} + \mathbf{x}_{2i}'\boldsymbol{\theta}_0)\mathbf{x}_{2i}\mathbf{x}_{2i}', \qquad \bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) = -\mathbb{E}[(2y-1)\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta})\mathbf{x}_2\mathbf{x}_2'],$$

we obtain the decomposition

$$\tilde{\mathbf{H}}_n^{\mathtt{MS}} = \check{\mathbf{H}}_n^{\mathtt{MS}} + \mathbf{R}_n^{\mathtt{MS}} + \mathbf{S}_n^{\mathtt{MS}},$$

where

$$\mathbf{R}_n^{\mathtt{MS}} = \tilde{\mathbf{H}}_n^{\mathtt{MS}} - \check{\mathbf{H}}_n^{\mathtt{MS}} - \bar{\mathbf{H}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) + \bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0), \qquad \mathbf{S}_n^{\mathtt{MS}} = \bar{\mathbf{H}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) - \bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0).$$

The proof will be completed by showing that $\check{\mathbf{H}}_n^{\mathtt{MS}} \to_{\mathbb{P}} \mathbf{H}^{\mathtt{MS}}, \mathbf{R}_n^{\mathtt{MS}} = o_{\mathbb{P}}(1)$, and $\mathbf{S}_n^{\mathtt{MS}} = o_{\mathbb{P}}(1)$.

First, using the dominated convergence theorem and $\int_{\mathbb{R}} u\dot{K}(u)du = -1$, we have

$$\begin{aligned}
\bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) &= -\mathbb{E}[(2y-1)\dot{K}_n\left(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0\right)\mathbf{x}_2\mathbf{x}_2'] \\
&= -\mathbb{E}\left[\int_{\mathbb{R}} \frac{1 - 2F_{u|x_1,\mathbf{x}_2}(-uh_n|uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}{h_n} f_{x_1|\mathbf{x}_2}(uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\dot{K}(u)du\mathbf{x}_2\mathbf{x}_2'\right] \\
&\to 2\mathbb{E}[F_0^{(1,1)}(\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'] \int_{\mathbb{R}} u\dot{K}(u)du \\
&= 2\mathbb{E}[f_{u|x_1,\mathbf{x}_2}(0| - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'] = \mathbf{H}^{\mathtt{MS}}.
\end{aligned}$$

Moreover, $\check{\mathbf{H}}_n^{\mathtt{MS}} - \bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) = o_{\mathbb{P}}(1)$ because each element has mean zero and a variance that is bounded by a constant multiple of

$$\frac{\mathbb{E}[\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)^2]}{n} = O\left(\frac{1}{nh_n^3}\right) = o(1).$$

As a consequence, $\check{\mathbf{H}}_n^{\mathtt{MS}} \to_{\mathbb{P}} \mathbf{H}^{\mathtt{MS}}$.

Next, $\mathbf{R}_n^{\mathtt{MS}} = o_{\mathbb{P}}(1/\sqrt{nh_n^3}) = o_{\mathbb{P}}(1)$ follows from Pollard (1989, Theorem 4.2) if it can be shown that, for every $C > 0$,

$$h_n^3 \mathbb{E}\left[\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \leq Cn^{-1/3}} |\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}) - \dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)|^2 \|\mathbf{x}_2\|^4\right] = o(1).$$

Defining $\bar{K}_n(u) = \bar{K}(u/h_n)/h_n$, we have, by Condition K(iii),

$$\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \leq Cn^{-1/3}} |\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}) - \dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)| \leq \frac{C}{n^{1/3}h_n^2} \bar{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)\|\mathbf{x}_2\|,$$

and therefore, using $nh_n^3 \to \infty$,

$$\begin{aligned} & h_n^3 \mathbb{E}\left[\sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \leq Cn^{-1/3}} |\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}) - \dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)|^2 \|\mathbf{x}_2\|^4\right] \\ \leq\ & \frac{C^2}{n^{2/3}h_n} \mathbb{E}[\bar{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)^2 \|\mathbf{x}_2\|^6] = O\left(\frac{1}{n^{2/3}h_n^2}\right) = o(1). \end{aligned}$$

Finally, defining

$$\begin{aligned} \xi_n(u, \boldsymbol{\delta}, \mathbf{x}_2) \ =\ & \frac{1 - 2F_{u|x_1,\mathbf{x}_2}(-uh_n + \mathbf{x}_2'\boldsymbol{\delta} | uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0 - \mathbf{x}_2'\boldsymbol{\delta}, \mathbf{x}_2)}{h_n} f_{x_1|\mathbf{x}_2}(uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0 - \mathbf{x}_2'\boldsymbol{\delta} | \mathbf{x}_2) \\ & - \frac{1 - 2F_{u|x_1,\mathbf{x}_2}(-uh_n | uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}{h_n} f_{x_1|\mathbf{x}_2}(uh_n - \mathbf{x}_2'\boldsymbol{\theta}_0 | \mathbf{x}_2), \end{aligned}$$

we have

$$\begin{aligned} \sup_{||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| \leq Cn^{-1/3}} \left\|\bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) - \bar{\mathbf{H}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)\right\| &= \sup_{||\boldsymbol{\delta}|| \leq Cn^{-1/3}} \left\|\mathbb{E}\left[\int_{\mathbb{R}} \xi_n(u, \boldsymbol{\delta}, \mathbf{x}_2)\dot{K}(u) du \mathbf{x}_2\mathbf{x}_2'\right]\right\| \\ &\leq \mathbb{E}\left[\left\{\int_{\mathbb{R}} \sup_{||\boldsymbol{\delta}|| \leq Cn^{-1/3}} |\xi_n(u, \boldsymbol{\delta}, \mathbf{x}_2)||\dot{K}(u)| du\right\} ||\mathbf{x}_2||^2\right] \\ &\to 0 \end{aligned}$$

for any $C > 0$, where the last line uses the dominated convergence theorem.

A.2.2.2 Part (ii) [Approximate MSE]

It was shown in the proof of part (i) that $R_{n,kl}^{\mathtt{MS}} = o_{\mathbb{P}}(1/\sqrt{nh_n^3})$. Also, Taylor's theorem and Condition K(ii) can be used to show that for any $C > 0$, we have, uniformly in $||\boldsymbol{\delta}_n|| \leq C/\sqrt[3]{n}$,

$$\begin{aligned} \bar{H}_{n,kl}^{\mathtt{MS}}(\boldsymbol{\theta}_0 + \boldsymbol{\delta}_n) \ =\ & H_{kl}^{\mathtt{MS}} + h_n^2 \mathbb{E}[\{F_0^{(1,3)}(\mathbf{x}_2) + F_0^{(2,2)}(\mathbf{x}_2) + F_0^{(3,1)}(\mathbf{x}_2)/3\} x_{2,k} x_{2,l}] \int_{\mathbb{R}} u^3 \dot{K}(u) du \\ & + \{4\mathbb{E}[F_0^{(1,2)}(\mathbf{x}_2) x_{2,k} x_{2,l} \mathbf{x}_2] + 2\mathbb{E}[F_0^{(2,1)}(\mathbf{x}_2) x_{2,k} x_{2,l} \mathbf{x}_2]\}' \boldsymbol{\delta}_n + o(h_n^2), \end{aligned}$$

implying in particular that

$$S_{n,kl}^{\mathtt{MS}} = \{4\mathbb{E}[F_0^{(1,2)}(\mathbf{x}_2)x_{2,k}x_{2,l}\mathbf{x}_2] + 2\mathbb{E}[F_0^{(2,1)}(\mathbf{x}_2)x_{2,k}x_{2,l}\mathbf{x}_2]\}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{\mathbb{P}}(h_n^2).$$

As a consequence, $\tilde{H}_{n,kl}^{\mathtt{ND}} - \check{H}_{n,kl}^{\mathtt{ND}} = o_{\mathbb{P}}(h_n^2 + 1/\sqrt{nh_n^3}) + O_{\mathbb{P}}(1/\sqrt[3]{n})$, where the $O_{\mathbb{P}}(1/\sqrt[3]{n})$ term

$$\{4\mathbb{E}[F_0^{(1,2)}(\mathbf{x}_2)x_{2,k}x_{2,l}\mathbf{x}_2] + 2\mathbb{E}[F_0^{(2,1)}(\mathbf{x}_2)x_{2,k}x_{2,l}\mathbf{x}_2]\}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$

does not depend on h_n.

Next, we approximate the moments of $\check{H}_{n,kl}^{\mathtt{MS}}$. By the previous paragraph,

$$\mathbb{E}[\check{H}_{n,kl}^{\mathtt{MS}}] - H_{kl}^{\mathtt{MS}} = \bar{H}_{n,kl}^{\mathtt{MS}}(\boldsymbol{\theta}_0) - H_{kl}^{\mathtt{MS}} = h_n^2\mathsf{B}_{kl} + o(\epsilon_n^2),$$

where

$$\mathsf{B}_{kl} = \mathbb{E}[\{F_0^{(1,3)}(\mathbf{x}_2) + F_0^{(2,2)}(\mathbf{x}_2) + F_0^{(3,1)}(\mathbf{x}_2)/3\}x_{2,k}x_{2,l}]\int_{\mathbb{R}} u^3\dot{K}(u)du.$$

Also,

$$\begin{aligned}
\mathbb{V}[\check{H}_{n,kl}^{\mathtt{MS}}] &= \frac{1}{n}\mathbb{V}[(2y-1)\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)x_{2,k}x_{2,l}] = \frac{1}{n}\mathbb{V}[\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)x_{2,k}x_{2,l}] \\
&= \frac{1}{n}\mathbb{E}[\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)^2x_{2,k}^2x_{2,l}^2] + O\left(\frac{1}{n}\right) \\
&= \frac{1}{nh_n^3}\mathsf{V}_{kl} + o\left(\frac{1}{nh_n^3}\right),
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{V}_{kl} &= \lim_{n\to\infty} h_n^3\mathbb{E}[\dot{K}_n(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_0)^2x_{2,k}^2x_{2,l}^2] = \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)x_{2,k}^2x_{2,l}^2]\int_{\mathbb{R}}\dot{K}(u)^2du \\
&= 2\mathbb{E}[F_0^{(0,1)}(\mathbf{x}_2)x_{2,k}^2x_{2,l}^2]\int_{\mathbb{R}}\dot{K}(u)^2du.
\end{aligned}$$

A.2.3 Rule-of-Thumb Bandwidth Selection

We provide details on the rule-of-thumb (ROT) bandwidth selection rules used in the simulations reported below. To construct ROT bandwidths, we choose a reference model involving finite dimensional parameters and calculate/approximate the corresponding leading constants entering the approximate MSE of $\tilde{\mathbf{H}}_n^{\mathtt{MS}}$ and $\tilde{\mathbf{H}}_n^{\mathtt{ND}}$.

Specifically, we assume $u|\mathbf{x} \sim \mathcal{N}(0, \sigma_u^2(\mathbf{x}))$ and $x_1|\mathbf{x}_2 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, where we will specify some parametric specification on $\sigma_u^2(\mathbf{x}) = \sigma_u^2(x_1, \mathbf{x}_2)$. Then, in this reference model, $F_0^{(2,2)}(\mathbf{x}_2) = 0$,

$$F_0^{(1,3)}(\mathbf{x}_2) = -\frac{\phi(0)}{\sigma_u(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)\sigma_1^3}\phi\left(\frac{\mathbf{x}_2'\boldsymbol{\theta}_0 + \mu_1}{\sigma_1}\right)\left[\left(\frac{\mathbf{x}_2'\boldsymbol{\theta}_0 + \mu_1}{\sigma_1}\right)^2 - 1\right],$$

and

$$F_0^{(3,1)}(\mathbf{x}_2) = \frac{\phi(0)}{\sigma_u^3(\mathbf{x})\sigma_1}\phi\left(\frac{\mathbf{x}_2'\boldsymbol{\theta}_0 + \mu_1}{\sigma_1}\right)[1 - \ddot{\sigma}_u(\mathbf{x})\sigma_u(\mathbf{x}) + 2\dot{\sigma}_u(\mathbf{x})^2]\Big|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0},$$

where ϕ is the standard normal density and where $\dot{\sigma}_u(\mathbf{x}) = \partial\sigma_u(\mathbf{x})/\partial x_1$ and $\ddot{\sigma}_u(\mathbf{x}) = \partial^2\sigma_u(\mathbf{x})/\partial x_1^2$.

A.2.3.1 Plug-in Estimator $\tilde{\mathbf{H}}_n^{\texttt{MS}}$

Given our reference model, natural estimators of the bias constants

$$\mathsf{B}_{kl} = \mathbb{E}[\{F_0^{(1,3)}(\mathbf{x}_2) + F_0^{(3,1)}(\mathbf{x}_2)/3\}x_{2,k}x_{2,l}]\int_{\mathbb{R}} u^3\dot{K}(u)du$$

are

$$\left[\frac{1}{n}\sum_{i=1}^{n}\{\hat{F}_n^{(1,3)}(\mathbf{x}_{2i}) + \hat{F}_n^{(3,1)}(\mathbf{x}_{2i})/3\}\mathbf{e}_k'\mathbf{x}_{2i}\mathbf{e}_l'\mathbf{x}_{2i}\right]\int_{\mathbb{R}} u^3\dot{K}(u)du,$$

where $\hat{F}_n^{(1,3)}$ and $\hat{F}_n^{(3,1)}$ are constructed using maximum likelihood for the parametric reference model (i.e., heteroskedastic Probit) together with a flexible parametric specification $\sigma_u^2(\mathbf{x}) = \boldsymbol{\gamma}'\mathbf{p}(\mathbf{x})$ for $\sigma_u^2(\mathbf{x})$, with $\mathbf{p}(\mathbf{x})$ denoting a polynomial expansion.

Similarly, natural estimators of the variance constants

$$\mathsf{V} = 2\mathbb{E}[F_0^{(0,1)}(\mathbf{x}_2)x_{2,k}^2x_{2,l}^2]\int_{\mathbb{R}}\dot{K}(u)^2du,$$

are given by

$$\hat{\mathsf{V}}_n = 2\left[\frac{1}{n}\sum_{i=1}^{n}\hat{F}_n^{(0,1)}(\mathbf{x}_{2i})(\mathbf{e}_k'\mathbf{x}_{2i})^2(\mathbf{e}_l'\mathbf{x}_{2i})^2\right]\int_{\mathbb{R}}\dot{K}(u)^2du.$$

A.2.3.2 Numerical Differentiation Estimator $\tilde{\mathbf{H}}_n^{\texttt{ND}}$

In our reference model, the bias constants are of the form

$$\mathsf{B}_{kl} = -\mathbb{E}[\{F_0^{(1,3)}(\mathbf{x}_2) + F_0^{(3,1)}(\mathbf{x}_2)/3\}\{x_{2,k}^3x_{2,l} + x_{2,k}x_{2,l}^3\}],$$

natural estimators of which are given by

$$-\frac{1}{n}\sum_{i=1}^{n}\{\hat{F}_n^{(1,3)}(\mathbf{x}_{2i}) + \hat{F}_n^{(3,1)}(\mathbf{x}_{2i})/3\}\{(\mathbf{e}_k'\mathbf{x}_{2i})^3(\mathbf{e}_l'\mathbf{x}_{2i}) + (\mathbf{e}_k'\mathbf{x}_{2i})(\mathbf{e}_l'\mathbf{x}_{2i})^3\}.$$

Similarly, natural estimators of the variance constants

$$\mathsf{V}_{kl} = \{2\mathcal{B}_0(\mathbf{e}_k) + 2\mathcal{B}_0(\mathbf{e}_l) - \mathcal{B}_0(\mathbf{e}_k + \mathbf{e}_l) - \mathcal{B}_0(\mathbf{e}_k - \mathbf{e}_l)\}/16, \qquad \mathcal{B}_0(\mathbf{s}) = 2\mathbb{E}[F_0^{(0,1)}(\mathbf{x}_2)|\mathbf{x}_2'\mathbf{s}|],$$

are given by

$$\frac{1}{8n}\sum_{i=1}^{n}\{2|\mathbf{e}_k'\mathbf{x}_{2i}| + 2|\mathbf{e}_l'\mathbf{x}_{2i}| - |(\mathbf{e}_k + \mathbf{e}_l)'\mathbf{x}_{2i}| - |(\mathbf{e}_k - \mathbf{e}_l)'\mathbf{x}_{2i}|\}\hat{F}_n^{(0,1)}(\mathbf{x}_{2i}).$$

A.3 Example: Panel Maximum Score

To state sufficient conditions for Condition CRA$_0$ in this example, define

$$\eta^{\texttt{PMS}}(\mathbf{x}_2) = \left\{\frac{\partial}{\partial x_1}\mathbb{E}(y|x_1, \mathbf{x}_2)\right\} f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)\bigg|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0}.$$

Condition PMS For some $\delta > 0$, the following are satisfied:

(i) For every $u \in \mathbb{R}$, $0 < F_{u_1|\mathbf{X}_1,\mathbf{X}_2,\alpha}(u_1|\mathbf{X}_1, \mathbf{X}_2, \alpha) = F_{u_2|\mathbf{X}_1,\mathbf{X}_2,\alpha}(u_2|\mathbf{X}_1, \mathbf{X}_2, \alpha) < 1$ almost surely. Also, $\mathbb{E}[y|x_1, \mathbf{x}_2]$ is continuously differentiable in x_1 with bounded derivative, and $\mathbb{E}[y^2|x_1, \mathbf{x}_2]$ is continuous in x_1.

(ii) The support of $\mathbf{x}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+1}$, $\mathbb{E}[\|\mathbf{x}_2\|^2] < \infty$, and conditional on $\mathbf{x}_2$, x_1 has everywhere positive Lebesgue density. Also, $F_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)$ is continuously differentiable in x_1 with bounded derivative.

(iii) $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

(iv) $M^{\texttt{PMS}}(\boldsymbol{\theta}) = \mathbb{E}[m^{\texttt{PMS}}(\mathbf{z}, \boldsymbol{\theta})]$ is twice continuously differentiable in $\boldsymbol{\theta}$ on $\boldsymbol{\Theta}_0^\delta$ and

$$\mathbf{H}^{\texttt{PMS}} = \mathbb{E}[\eta^{\texttt{PMS}}(\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2']$$

is positive definite.

Letting

$$\psi^{\texttt{PMS}}(\mathbf{x}_2) = \mathbb{E}(y^2|x_1, \mathbf{x}_2) f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)\big|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0}$$

and proceeding as in the proof of Corollary MS, the following result is obtained.

Corollary PMS *Suppose Condition PMS is satisfied. Then Condition CRA is satisfied with* $q_n = 1$, $\mathbf{H}_0 = \mathbf{H}^{\texttt{PMS}}$, *and* $\mathcal{C}_0 = \mathcal{C}^{\texttt{PMS}}$, *where*

$$\mathcal{C}^{\texttt{PMS}}(\mathbf{s}, \mathbf{t}) = \mathbb{E}[\psi^{\texttt{PMS}}(\mathbf{x}_2) \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} \mathbb{1}(\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t}))].$$

The case-specific estimator $\tilde{\mathbf{H}}_n^{\texttt{PMS}}$ of $\mathbf{H}^{\texttt{PMS}}$ admits a counterpart of Lemma MS, but for brevity we omit a precise statement.

A.4 Example: Conditional Maximum Score

To state sufficient conditions for Condition CRA in this example, let $\mathcal{X}$ denote the support of $\mathbf{x} = (x_1, \mathbf{x}_2')'$ and for $\delta > 0$, let $\mathcal{W}^\delta = \{\mathbf{w} \in \mathbb{R}^d : ||\mathbf{w}|| \leq \delta\}$. Also, define

$$\mu^{\texttt{CMS}}(\mathbf{w}; \boldsymbol{\theta}) = \mathbb{E}\left[y \mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0) \middle| \mathbf{w} \right] f_{\mathbf{w}}(\mathbf{w}),$$

$$\dot{\mu}^{\texttt{CMS}}(\mathbf{w}; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mu^{\texttt{CMS}}(\mathbf{w}; \boldsymbol{\theta}) = \mathbb{E}\left[\{\mathbb{E}(y|x_1, \mathbf{x}_2, \mathbf{w}) f_{x_1|\mathbf{x}_2,\mathbf{w}}(x_1|\mathbf{x}_2, \mathbf{w})\}\big|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}} \mathbf{x}_2 \middle| \mathbf{w} \right] f_{\mathbf{w}}(\mathbf{w}),$$

$$\ddot{\mu}^{\texttt{CMS}}(\mathbf{w}; \boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mu^{\texttt{CMS}}(\mathbf{w}; \boldsymbol{\theta}),$$

and

$$\eta^{\texttt{CMS}}(\mathbf{x}_2) = \left\{ \frac{\partial}{\partial x_1} \mathbb{E}(y|x_1, \mathbf{x}_2, \mathbf{w}) \right\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(x_1|\mathbf{x}_2, \mathbf{w}) \bigg|_{x_1 = -\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{w} = \mathbf{0}}.$$

Condition CMS For some $\delta > 0$ and $P \geq 1$, the following are satisfied:

(i) For some strictly increasing F,

$$\mathbb{P}(Y_t = 1|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \alpha, Y_0, \dots, Y_{t-1}) = F[X_{1t} + (\mathbf{X}_{2t}', Y_{t-1})\boldsymbol{\theta}_0 + \alpha], \qquad t = 1, 2, 3.$$

Also, on $\mathcal{X} \times \mathcal{W}^\delta$, $\mathbb{E}(y|x_1, \mathbf{x}_2, \mathbf{w})$ is differentiable in x_1, $\partial \mathbb{E}(y|x_1, \mathbf{x}_2, \mathbf{w})/\partial x_1$ is bounded and continuous in $(x_1, \mathbf{w})$, and $\mathbb{E}(y^2|x_1, \mathbf{x}_2, \mathbf{w})$ is positive and continuous in $(x_1, \mathbf{w})$.

(ii) $\mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}]$ is bounded on $\mathcal{W}^\delta$ and for every $\mathbf{w} \in \mathcal{W}^\delta$, the support of $\mathbf{x}$ given $\mathbf{w}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+1}$. Also, on $\mathcal{X} \times \mathcal{W}^\delta$, $f_{x_1|\mathbf{x}_2,\mathbf{w}}(x_1|\mathbf{x}_2,\mathbf{w})$ is positive, bounded, and continuous in $(x_1, \mathbf{w})$ and $f_\mathbf{w}(\mathbf{w})$ is positive and continuous in $\mathbf{w}$.

(iii) $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

(iv) $\mu^{\texttt{CMS}}(\mathbf{w};\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$ on $\boldsymbol{\Theta}_0^\delta$ with bounded derivatives, $\mu^{\texttt{CMS}}(\mathbf{w};\boldsymbol{\theta})$ is uniformly (in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$) continuous in $\mathbf{w}$ at $\mathbf{0}$, $\dot{\mu}^{\texttt{CMS}}(\mathbf{w};\boldsymbol{\theta}_0)$ is P times continuously differentiable in $\mathbf{w}$ on $\mathcal{W}^\delta$, $\ddot{\mu}^{\texttt{CMS}}(\mathbf{w};\boldsymbol{\theta})$ is uniformly (in $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\delta$) continuous in $\mathbf{w}$ at $\mathbf{0}$, and

$$\mathbf{H}^{\texttt{CMS}} = \mathbb{E}\left[\eta^{\texttt{CMS}}(\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' \middle| \mathbf{w}\right] f_\mathbf{w}(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{0}}$$

is positive definite.

(v) κ is bounded, of order P, and supported on $[-1,1]^d$. Also, $nb_n^d \to \infty$ and $nb_n^{d+3P} \to 0$.

Let

$$\psi^{\texttt{CMS}}(\mathbf{x}_2) = \mathbb{E}(y^2|x_1,\mathbf{x}_2,\mathbf{w}) f_{x_1|\mathbf{x}_2,\mathbf{w}}(x_1|\mathbf{x}_2,\mathbf{w})\Big|_{x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{w}=\mathbf{0}}.$$

Corollary CMS *Suppose Condition CMS is satisfied. Then Condition CRA is satisfied with* $q_n = b_n^d$, $\mathbf{H}_0 = \mathbf{H}^{\texttt{CMS}}$, *and* $\mathcal{C}_0 = \mathcal{C}^{\texttt{CMS}}$, *where*

$$\mathcal{C}^{\texttt{CMS}}(\mathbf{s},\mathbf{t}) = \mathbb{E}\left[\psi^{\texttt{CMS}}(\mathbf{x}_2)\min\{|\mathbf{x}_2'\mathbf{s}|,|\mathbf{x}_2'\mathbf{t}|\}\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\} \middle| \mathbf{w}\right] f_\mathbf{w}(\mathbf{w})\Big|_{\mathbf{w}=\mathbf{0}} \cdot \int_{\mathbb{R}^d} \kappa(\mathbf{v})^2 d\mathbf{v},$$

The case-specific estimator $\tilde{\mathbf{H}}_n^{\texttt{CMS}}$ of $\mathbf{H}^{\texttt{CMS}}$ admits a counterpart of Lemma MS, but for brevity we omit a precise statement.

A.4.1 Proof of Corollary CMS

Condition CRA(i). Because κ_n does not depend on $\boldsymbol{\theta}$, uniform manageability can be established by proceeding as in the maximum score example.

Also, $\bar{m}_n(\mathbf{z}) = |\kappa_n(\mathbf{w})|$ satisfies $q_n\mathbb{E}[\bar{m}_n(\mathbf{z})^2]=b_n^d O(1/b_n^d) = O(1)$.

Next, using the representations

$$M_n(\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \mu(\mathbf{v}b_n;\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v} \qquad \text{and} \qquad M_0(\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \mu(\mathbf{0};\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v},$$

we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})| \le \left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta},||\mathbf{w}||\le b_n} |\mu(\mathbf{w};\boldsymbol{\theta}) - \mu(\mathbf{0};\boldsymbol{\theta})|\right\} \int_{\mathbb{R}^d} |\kappa(\mathbf{v})|d\mathbf{v} = o(1),$$

where the equality uses uniform (in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$) continuity of $\mu(\mathbf{w};\boldsymbol{\theta})$ at $\mathbf{w} = \mathbf{0}$.

Finally, well-separatedness follows from compactness of $\boldsymbol{\Theta}$, continuity of $M_0(\boldsymbol{\theta}) = \mu(0;\boldsymbol{\theta})$ in $\boldsymbol{\theta}$, and the fact (shown by Honoré and Kyriazidou (2000, Lemmas 6 and 7)) that $\boldsymbol{\theta}_0$ is the unique maximizer of $M_0(\boldsymbol{\theta})$.

Condition CRA(ii). We have

$$\frac{\partial}{\partial\boldsymbol{\theta}} M_n(\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \dot{\mu}(\mathbf{v}b_n;\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v}, \qquad \frac{\partial}{\partial\boldsymbol{\theta}} M_0(\boldsymbol{\theta}) = \int_{\mathbb{R}^d} \dot{\mu}(\mathbf{0};\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v},$$

and

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_n(\boldsymbol{\theta}) = \int_{\mathbb{R}^d}\ddot{\mu}(\mathbf{v}b_n;\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v}, \qquad \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_0(\boldsymbol{\theta}) = \int_{\mathbb{R}^d}\ddot{\mu}(\mathbf{0};\boldsymbol{\theta})\kappa(\mathbf{v})d\mathbf{v},$$

where, using uniform (in $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta$) continuity of $\ddot{\mu}(\mathbf{w};\boldsymbol{\theta})$ at $\mathbf{w}=\mathbf{0}$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta})-M_0(\boldsymbol{\theta})]\right\| \le \left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta,||\mathbf{w}||\le b_n}|\ddot{\mu}(\mathbf{w};\boldsymbol{\theta})-\ddot{\mu}(\mathbf{0};\boldsymbol{\theta})|\right\}\int_{\mathbb{R}^d}|\kappa(\mathbf{v})|d\mathbf{v} = o(1).$$

Also, by a standard bias calculation for kernel estimators,

$$\frac{\partial}{\partial\boldsymbol{\theta}}M_n(\boldsymbol{\theta}_0) = \int_{\mathbb{R}^d}\dot{\mu}(\mathbf{v}b_n;\boldsymbol{\theta}_0)\kappa(\mathbf{v})d\mathbf{v} = \dot{\mu}(\mathbf{0};\boldsymbol{\theta}_0)+O(b_n^P),$$

where it follows from Honoré and Kyriazidou (2000) that $\dot{\mu}(\mathbf{0};\boldsymbol{\theta}_0) = \partial M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} = \mathbf{0}$. As a consequence, $r_n||\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}|| = O(\sqrt[3]{nb_n^{d+3P}}) = o(1)$.

Finally, $\mathbf{H}_0 = -\ddot{\mu}(\mathbf{0};\boldsymbol{\theta}_0) = \mathbf{H}^{\mathtt{CMS}}$ is positive definite by assumption.

Condition CRA(iii). Because κ_n does not depend on $\boldsymbol{\theta}$, uniform manageability can be established by proceeding as in the maximum score example. For this example,

$$\bar{d}_n^\delta(\mathbf{z}) = \left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta}_0 < -x_1 \le \mathbf{x}_2'\boldsymbol{\theta}) + \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\delta}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta} < -x_1 \le \mathbf{x}_2'\boldsymbol{\theta}_0)\right]|\kappa_n(\mathbf{w})|.$$

By change of variables and using boundedness of $f_{x_1|\mathbf{x}_2,\mathbf{w}}(x_1|\mathbf{x}_2,\mathbf{w})$, we have, uniformly in δ,

$$b_n^d\mathbb{E}[\bar{d}_n^\delta(\mathbf{z})^2/\delta] = b_n^d O(\mathbb{E}[\|\mathbf{x}_2\|\kappa_n(\mathbf{w})^2]) = O(1).$$

As a consequence, $q_n\sup_{0<\delta'\le\delta}\mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$.

Condition CRA(iv). Using $\bar{d}_n^\delta(\mathbf{z})^4 \le 8d_n^\delta(\mathbf{z})|\kappa_n(\mathbf{w})|^3$, it follows from calculations similar to those above that $q_n^3r_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = O(r_n^{-1}\delta_n) = o(1)$.

As in the maximum score example, $\mathcal{C}^{\mathtt{CMS}}(\mathbf{s},\mathbf{s})+\mathcal{C}^{\mathtt{CMS}}(\mathbf{t},\mathbf{t})-2\mathcal{C}^{\mathtt{CMS}}(\mathbf{s},\mathbf{t}) > 0$.

Finally, the representation

$$\begin{aligned}&\{m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s})-m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta})\}\{m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t})-m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta})\}\\ &= y^2[\mathbb{1}\{\delta_n\min(\mathbf{x}_2'\mathbf{s},\mathbf{x}_2'\mathbf{t}) \ge -x_1-\mathbf{x}_2'\boldsymbol{\theta} > 0\} + \mathbb{1}\{\delta_n\max(\mathbf{x}_2'\mathbf{s},\mathbf{x}_2'\mathbf{t}) < -x_1-\mathbf{x}_2'\boldsymbol{\theta} \le 0\}]\kappa_n(\mathbf{w})^2\end{aligned}$$

can be used to show that, uniformly in $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}$,

$$\frac{q_n}{\delta_n}\mathbb{E}[\{m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s})-m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta})\}\{m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t})-m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta})\}] = \mathcal{C}^{\mathtt{CMS}}(\mathbf{s},\mathbf{t})+o(1).$$

Condition CRA(v). The first condition follows from $q_n\bar{d}_n^\delta(\mathbf{z}) \le \sup_{\mathbf{v}\in\mathbb{R}^d}|\kappa(\mathbf{v})|$. The second condition follows from the calculation similar to the covariance kernel calculation.

A.5 Example: Empirical Risk Minimization

In this example, we follow Mohammadi and van de Geer (2005, Theorem 1) when stating primitive conditions. Let F denote the distribution function of x and let $P(x) = \mathbb{P}[y=1|x]$.

Condition ERM The following are satisfied:

(i) $P(0) < 1/2$ and P admits a continuous derivative p in a neighborhood of each element of $\boldsymbol{\theta}_0$.

(ii) F is absolutely continuous and its Lebesgue density f is continuously differentiable in a neighborhood of each element of $\boldsymbol{\theta}_0$.

(iii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

(iv) $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}, \cdots, \theta_{0,d})'$ is the unique minimizer of $\mathbb{P}[h_{\boldsymbol{\theta}}(x) \neq y]$ and $p(\theta_{0,\ell})f(\theta_{0,\ell}) \neq 0$ for $\ell \in \{1, \cdots, d\}$.

Corollary ERM *Suppose Condition ERM is satisfied. Then Condition CRA is satisfied with* $q_n = 1$, $\mathbf{H}_0 = \mathbf{H}^{\texttt{ERM}}$, *and* $\mathcal{C}_0 = \mathcal{C}^{\texttt{ERM}}$, *where*

$$\mathbf{H}^{\texttt{ERM}} = 2\begin{pmatrix} p(\theta_{0,1})f(\theta_{0,1}) & 0 & \cdots & 0 \\ 0 & -p(\theta_{0,2})f(\theta_{0,2}) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & (-1)^{d+1}p(\theta_{0,d})f(\theta_{0,d}) \end{pmatrix}$$

and, for $\mathbf{s} = (s_1, \cdots, s_d)'$ *and* $\mathbf{t} = (t_1, \cdots, t_d)'$,

$$\mathcal{C}^{\texttt{ERM}}(\mathbf{s}, \mathbf{t}) = \sum_{\ell=1}^{d} f(\theta_{0,\ell}) \min\{|s_\ell|, |t_\ell|\} \mathbb{1}\{\operatorname{sgn}(s_\ell) = \operatorname{sgn}(t_\ell)\}.$$

A case-specific (plug-in) estimator of $\mathbf{H}^{\texttt{ERM}}$ is given by the diagonal matrix $\tilde{\mathbf{H}}_n^{\texttt{ERM}}$ with diagonal elements

$$\tilde{H}_{n,\ell\ell}^{\texttt{ERM}} = (-1)^{\ell+1} 2\hat{p}_n(\hat{\theta}_{n,\ell}^{\texttt{ERM}})\hat{f}_n(\hat{\theta}_{n,\ell}^{\texttt{ERM}}), \qquad \ell = 1, \ldots, d,$$

where $\hat{p}_n$ and $\hat{f}_n$ are some nonparametric estimators of p and f. This estimator is consistent whenever its ingredients $\hat{p}_n$ and $\hat{f}_n$ are.

A.5.1 Proof of Corollary ERM

By Lemma A.11, it suffices to verify that Condition CRA_0 is satisfied.

Condition CRA_0(i). Manageability of $\mathcal{M}_0$ follows from $\{\mathbb{1}(h_{\boldsymbol{\theta}}(x) \neq y) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ forming a VC subgraph class. Also, the envelope is bounded by 1. Finally, $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) < M_0(\boldsymbol{\theta}_0)$ for every $\delta > 0$ because $\boldsymbol{\Theta}$ is compact, M_0 is continuous, and $\boldsymbol{\theta}_0$ is the unique maximizer of $M_0(\boldsymbol{\theta})$.

Condition CRA_0(ii). By assumption, $\boldsymbol{\theta}_0$ belongs to the interior of $\boldsymbol{\Theta}_0$. Mohammadi and van de Geer (2005) show that, for odd ℓ,

$$\frac{\partial^2}{\partial \theta_\ell^2} \mathbb{P}(h_{\boldsymbol{\theta}}(x) \neq y) = 2p(\theta_\ell)f(\theta_\ell) + (2P(\theta_\ell) - 1)\frac{d}{d\theta} f(\theta_\ell),$$

and that a similar formula holds for even ℓ as well. In particular, M_0 is twice continuous differentiability on $\boldsymbol{\Theta}_0^\delta$. Finally, positive definiteness of $\mathbf{H}_0 = \mathbf{H}^{\texttt{ERM}}$ is established in Mohammadi and van de Geer (2005).

Condition CRA$_0$(iii). This condition corresponds to the first part of (vii) in Theorem 7 of Mohammadi and van de Geer (2005).

Condition CRA$_0$(iv). Since $d_0^\delta(\mathbf{z})^4 = d_0^\delta(\mathbf{z})$, $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^4] = O(\delta_n)$, which implies the first condition. For the second part, Mohammadi and van de Geer (2005) show that

$$\mathcal{C}_0(\mathbf{s},\mathbf{t}) = \sum_{\ell=1}^{d} f(\theta_{0,\ell})[\min\{s_\ell,t_\ell\}\mathbb{1}(s_\ell>0,t_\ell>0) - \max\{s_\ell,t_\ell\}\mathbb{1}(s_\ell<0,t_\ell<0)].$$

Using the representations

$$m_0(1,x,\boldsymbol{\theta}) = -\sum_{\ell=0}^{\lfloor d/2\rfloor} \mathbb{1}\left(x\in[\theta_{2\ell},\theta_{2\ell+1})\right), \qquad m_0(-1,x,\boldsymbol{\theta}) = -\sum_{\ell=1}^{\lfloor (d+1)/2\rfloor} \mathbb{1}\left(x\in[\theta_{2\ell-1},\theta_{2\ell})\right),$$

it can be shown that, for $\boldsymbol{\theta}$ in the interior of $\boldsymbol{\Theta}$ and for δ_n small enough,

$$\begin{aligned}&\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s}) - m_0(\mathbf{z},\boldsymbol{\theta})\}\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t}) - m_0(\mathbf{z},\boldsymbol{\theta})\}\\ &= \sum_{\ell=1}^{d} \mathbb{1}(x\in[\theta_\ell+\delta_n\max\{s_\ell,t_\ell\},\theta_\ell))\mathbb{1}(s_\ell<0,t_\ell<0) + \mathbb{1}(x\in[\theta_\ell,\theta_\ell+\delta_n\min\{s_\ell,t_\ell\}))\mathbb{1}(s_\ell>0,t_\ell>0).\end{aligned}$$

As a consequence,

$$\begin{aligned}&\frac{1}{\delta_n}\mathbb{E}[\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s}) - m_0(\mathbf{z},\boldsymbol{\theta})\}\{m_0(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t}) - m_0(\mathbf{z},\boldsymbol{\theta})\}]\\ =& \sum_{\ell=1}^{d} f(\theta_\ell)[-\max\{s_\ell,t_\ell\}\mathbb{1}(s_\ell<0,t_\ell<0) + \min\{s_\ell,t_\ell\}\mathbb{1}(s_\ell>0,t_\ell>0)] + o(1)\\ =& \sum_{\ell=1}^{d} f(\theta_{0,\ell})[-\max\{s_\ell,t_\ell\}\mathbb{1}(s_\ell<0,t_\ell<0) + \min\{s_\ell,t_\ell\}\mathbb{1}(s_\ell>0,t_\ell>0)] + o(1)\end{aligned}$$

uniformly in $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}$.

Condition CRA$_0$(v). The condition in display is identical to the second part of (vii) in Theorem 7 of Mohammadi and van de Geer (2005). The second assumption corresponds to (vi) in Theorem 7 of Mohammadi and van de Geer (2005).

References

Abrevaya, J., and J. Huang (2005): "On the Bootstrap of the Maximum Score Estimator," *Econometrica*, 73(4), 1175–1204.

Honoré, B. E., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68(4), 839–874.

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18(1), 191–219.

Mohammadi, L., and S. van de Geer (2005): "Asymptotics in Empirical Risk Minimization," *Journal of Machine Learning Research*, 6(Dec), 2027–2047.

Pollard, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4(4), 341–354.