# Bootstrap-Based Inference for Cube Root Consistent Estimators*

Matias D. Cattaneo[†] Michael Jansson[‡] Kenichi Nagasawa[§]

May 30, 2018

**Abstract**

This note proposes a consistent bootstrap-based distributional approximation for cube root consistent estimators such as the maximum score estimator of Manski (1975) and the conditional maximum score estimator of Honoré and Kyriazidou (2000). For estimators of this kind, the standard nonparametric bootstrap is inconsistent. Our method restores consistency of the nonparametric bootstrap by altering the shape of the criterion function defining the estimator whose distribution we seek to approximate. This modification leads to a generic and easy-to-implement resampling method for inference that is conceptually distinct from other available distributional approximations, and can also be used in other related settings such as for the isotonic density estimator of Grenander (1956).

*Keywords:* cube root asymptotics, bootstrapping, maximum score estimation.

**JEL**: C12, C14, C21.

*For comments and suggestions, we are grateful to the handling coeditor, three referees, Mehmet Caner, Kei Hirano, Bo Honoré, Guido Kuersteiner, Andres Santos, and seminar participants at University of Aarhus, the 2017 Tinbergen Institute Conference, the 2017 California Econometrics Conference, the 2018 Workshop on Advanced Econometrics at University of Kansas, and the 2018 DAEiNA meeting. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grant SES-1459931, and Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1459967 and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78).

[†]Department of Economics and Department of Statistics, University of Michigan.

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

[§]Department of Economics, University of Michigan.

# 1 Introduction

In a seminal paper, Kim and Pollard (1990) studied a class of $M$-estimators exhibiting cube root asymptotics. These estimators not only have a non-standard rate of convergence, but also have the property that rather than being Gaussian their limiting distributions are of Chernoff (1964) type; that is, the limiting distribution is that of the maximizer of a Gaussian process. In fact, in leading examples of cube root consistent estimators such as the maximum score estimator of Manski (1975), the covariance kernel of the Gaussian process characterizing the limiting distribution depends on an infinite-dimensional nuisance parameter. From the perspective of inference, this feature of the limiting distribution represents a nontrivial complication relative to the conventional asymptotically normal case, where the limiting distribution is known up to the value of a finite-dimensional nuisance parameter (namely, the covariance matrix of the limiting distribution). In particular, the dependence of the limiting distribution on an infinite-dimensional nuisance parameter implies that resampling-based distributional approximations seem to offer the most attractive approach to inference in estimation problems exhibiting cube root asymptotics. The purpose of this note is to propose an easy-to-implement bootstrap-based distributional approximation applicable in such cases.

As does the familiar nonparametric bootstrap, the method proposed herein employs bootstrap samples of size $n$ from the empirical distribution function. But unlike the nonparametric bootstrap, which is inconsistent in general (e.g., Abrevaya and Huang, 2005; Léger and MacGibbon, 2006), our method offers a consistent distributional approximation for cube root consistent estimators and therefore has the advantage that it can be used to construct asymptotically valid inference procedures. Consistency is achieved by altering the shape of the criterion function defining the estimator whose distribution we seek to approximate. Heuristically, the method is designed to ensure that the bootstrap version of a certain empirical process has a mean resembling the large sample version of its population counterpart. The latter is quadratic in the problems we study, and known up to the value of a certain matrix. As a consequence, the only ingredient needed to implement the proposed "reshapement" of the objective function is a consistent estimator of the unknown matrix entering the quadratic mean of the empirical process. Such estimators turn out to be generically available and easy to compute.

The note proceeds as follows. Section 2 is heuristic in nature and serves the purpose of outlining the main idea underlying our approach in the $M$-estimation setting of Kim and Pollard (1990). Section 3 then makes the heuristics of Section 2 rigorous in a more general setting where the $M$-estimation problem is formed using a possibly $n$-varying (observation specific) objective function, as recently studied by Seo and Otsu (2018). Section 4 discusses two examples covered by our general results, namely the maximum score estimator of Manski (1975) and the conditional maximum score estimator of Honoré and Kyriazidou (2000). Further discussion of our results is provided in Section 5. Finally, all derivations and proofs have been collected in the supplemental appendix, where an extension to the case of the isotonic density estimator of Grenander (1956) is also given.

## 2 Heuristics

Suppose $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ is an estimand admitting the characterization

$$\boldsymbol{\theta}_0 = \operatorname*{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} M_0(\boldsymbol{\theta}), \qquad M_0(\boldsymbol{\theta}) = \mathbb{E}[m_0(\mathbf{z}, \boldsymbol{\theta})], \tag{1}$$

where $m_0$ is a known function, and where $\mathbf{z}$ is a random vector of which a random sample $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is available. Studying estimation problems of this kind for non-smooth $m_0$, Kim and Pollard (1990) gave conditions under which the $M$-estimator

$$\hat{\boldsymbol{\theta}}_n = \operatorname*{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \hat{M}_n(\boldsymbol{\theta}), \qquad \hat{M}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} m_0(\mathbf{z}_i, \boldsymbol{\theta}),$$

exhibits cube root asymptotics:

$$\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\}, \tag{2}$$

where $\rightsquigarrow$ denotes weak convergence, $\mathcal{Q}_0(\mathbf{s})$ is a quadratic form given by

$$\mathcal{Q}_0(\mathbf{s}) = -\frac{1}{2}\mathbf{s}'\mathbf{V}_0\mathbf{s}, \qquad \mathbf{V}_0 = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} M_0(\boldsymbol{\theta}_0),$$

and $\mathcal{G}_0$ is a non-degenerate zero-mean Gaussian process with $\mathcal{G}_0(0) = 0$.

Whereas the matrix $\mathbf{V}_0$ governing the shape of $\mathcal{Q}_0$ is finite-dimensional, the covariance kernel of

$\mathcal{G}_0$ in (2) typically involves infinite-dimensional unknown quantities. As a consequence, the limiting distribution of $\hat{\boldsymbol{\theta}}_n$ tends to be more difficult to approximate than Gaussian distributions, implying in turn that basing inference on $\hat{\boldsymbol{\theta}}_n$ is more challenging under cube root asymptotics than in the more familiar case where $\hat{\boldsymbol{\theta}}_n$ is ($\sqrt{n}$-consistent and) asymptotically normally distributed.

As a candidate method of approximating the distribution of $\hat{\boldsymbol{\theta}}_n$, consider the nonparametric bootstrap. To describe it, let $\mathbf{z}_{1,n}^*, \ldots, \mathbf{z}_{n,n}^*$ denote a random sample from the empirical distribution of $\mathbf{z}_1, \ldots \mathbf{z}_n$ and let the natural bootstrap analogue of $\hat{\boldsymbol{\theta}}_n$ be denoted by

$$\hat{\boldsymbol{\theta}}_n^* = \operatorname*{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \hat{M}_n^*(\boldsymbol{\theta}), \qquad \hat{M}_n^*(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} m_0(\mathbf{z}_{i,n}^*, \boldsymbol{\theta}).$$

Then the nonparametric bootstrap estimator of $\mathbb{P}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \leq \cdot]$ is given by $\mathbb{P}^*[\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n \leq \cdot]$, where $\mathbb{P}^*$ denotes a probability computed under the bootstrap distribution conditional on the data. As is well documented, however, this estimator is inconsistent under cube root asymptotics (e.g., Abrevaya and Huang, 2005; Léger and MacGibbon, 2006).

For the purpose of giving a heuristic, yet constructive, explanation of the inconsistency of the nonparametric bootstrap, it is helpful to recall that a proof of (2) can be based on the representation

$$\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \operatorname*{argmax}_{\mathbf{s} \in \mathbb{R}^d} \{Q_n(\mathbf{s}) + \hat{G}_n(\mathbf{s})\}, \tag{3}$$

where (for $\mathbf{s}$ such that $\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3} \in \boldsymbol{\Theta}$)

$$\hat{G}_n(\mathbf{s}) = n^{2/3}[\hat{M}_n(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_0(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) + M_0(\boldsymbol{\theta}_0)]$$

is a zero-mean random process, while

$$Q_n(\mathbf{s}) = n^{2/3}[M_0(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) - M_0(\boldsymbol{\theta}_0)]$$

is a non-random function that is correctly centered in the sense that $\operatorname{argmax}_{\mathbf{s} \in \mathbb{R}^d} Q_n(\mathbf{s}) = \mathbf{0}$. In cases where $m_0$ is non-smooth but $M_0$ is smooth, $Q_n$ and $\hat{G}_n$ are usually asymptotically quadratic and asymptotically Gaussian, respectively, in the sense that

$$Q_n(\mathbf{s}) \to \mathcal{Q}_0(\mathbf{s}) \tag{4}$$

and

$$\hat{G}_n(\mathbf{s}) \rightsquigarrow \mathcal{G}_0(\mathbf{s}). \tag{5}$$

Under regularity conditions ensuring among other things that the convergence in (4) and (5) is suitably uniform in $\mathbf{s}$, result (2) then follows from an application of a continuous mapping-type theorem for the argmax functional to the representation in (3).

Similarly to (3), the bootstrap analogue of $\hat{\boldsymbol{\theta}}_n$ admits a representation of the form

$$\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\hat{Q}_n(\mathbf{s}) + \hat{G}_n^*(\mathbf{s})\},$$

where (for $\mathbf{s}$ such that $\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3} \in \boldsymbol{\Theta}$)

$$\hat{G}_n^*(\mathbf{s}) = n^{2/3}[\hat{M}_n^*(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) - \hat{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n)]$$

and

$$\hat{Q}_n(\mathbf{s}) = n^{2/3}[\hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n)].$$

Under mild conditions, $\hat{G}_n^*$ satisfies the following bootstrap counterpart of (5):

$$\hat{G}_n^*(\mathbf{s}) \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0(\mathbf{s}), \tag{6}$$

where $\rightsquigarrow_{\mathbb{P}}$ denotes conditional weak convergence in probability (defined as van der Vaart and Wellner, 1996, Section 2.9). On the other hand, although $\hat{Q}_n$ is non-random under the bootstrap distribution and satisfies $\operatorname{argmax}_{\mathbf{s}\in\mathbb{R}^d} \hat{Q}_n(\mathbf{s}) = \mathbf{0}$, it turns out that $\hat{Q}_n(\mathbf{s}) \not\rightarrow_{\mathbb{P}} \mathcal{Q}_0(\mathbf{s})$ in general. In other words, the natural bootstrap counterpart of (4) typically fails and, as a partial consequence, so does the natural bootstrap counterpart of (2); that is, $\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \not\rightsquigarrow_{\mathbb{P}} \operatorname{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\}$ in general.

To the extent that the implied inconsistency of the bootstrap can be attributed to the fact that the shape of $\hat{Q}_n$ fails to replicate that of $Q_n$, it seems plausible that a consistent bootstrap-based distributional approximation can be obtained by basing the approximation on

$$\tilde{\boldsymbol{\theta}}_n^* = \operatorname*{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{M}_n^*(\boldsymbol{\theta}), \qquad \tilde{M}_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{m}_n(\mathbf{z}_{i,n}^*, \boldsymbol{\theta}),$$

where $\tilde{m}_n$ is a suitably "reshaped" version of $m_0$ satisfying two properties. First, $\tilde{Q}_n$ should be asymptotically quadratic, where $\tilde{Q}_n$ is the counterpart of $\hat{Q}_n$ associated with $\tilde{m}_n$ :

$$\tilde{Q}_n(\mathbf{s}) = n^{2/3}[\tilde{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) - \tilde{M}_n(\hat{\boldsymbol{\theta}}_n)], \qquad \tilde{M}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{m}_n(\mathbf{z}_i, \boldsymbol{\theta}).$$

Second, $\tilde{G}_n^*$ should be asymptotically equivalent to $\hat{G}_n^*$, where

$$\tilde{G}_n^*(\mathbf{s}) = n^{2/3}[\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) - \tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \tilde{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s}n^{-1/3}) + \tilde{M}_n(\hat{\boldsymbol{\theta}}_n)],$$

is the counterpart of $\hat{G}_n^*$ associated with $\tilde{m}_n$.

Accordingly, let

$$\tilde{m}_n(\mathbf{z}, \boldsymbol{\theta}) = m_0(\mathbf{z}, \boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n)'\tilde{\mathbf{V}}_n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n),$$

where $\tilde{\mathbf{V}}_n$ is an estimator of $\mathbf{V}_0$. Then

$$\sqrt[3]{n}(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d} \{\tilde{Q}_n(\mathbf{s}) + \tilde{G}_n^*(\mathbf{s})\},$$

where, by construction, $\tilde{G}_n^*(\mathbf{s}) = \hat{G}_n^*(\mathbf{s})$ and $\tilde{Q}_n(\mathbf{s}) = -\mathbf{s}'\tilde{\mathbf{V}}_n\mathbf{s}/2$. Because $\tilde{G}_n^* = \hat{G}_n^*$, $\tilde{G}_n^*(\mathbf{s}) \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0(\mathbf{s})$ whenever (6) holds. In addition, $\tilde{Q}_n(\mathbf{s}) \to_{\mathbb{P}} \mathcal{Q}_0(\mathbf{s})$ provided $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$. As a consequence, it seems plausible that $\mathbb{P}^*[\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n \leq \cdot]$ is a consistent estimator of $\mathbb{P}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \leq \cdot]$ if $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$.

## 3 Main Result

When making the heuristics of Section 2 precise, it is of interest to consider the more general situation where the estimator $\hat{\boldsymbol{\theta}}_n$ is an approximate maximizer (with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$) of

$$\hat{M}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} m_n(\mathbf{z}_i, \boldsymbol{\theta}),$$

where $m_n$ is a known function, and where $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a random sample of a random vector $\mathbf{z}$. This formulation of $\hat{M}_n$, which reduces to that of Section 2 when $m_n$ does not depend on $n$, is adopted in order to cover certain estimation problems where, rather than admitting a characterization of the form (1), the estimand $\boldsymbol{\theta}_0$ admits the characterization

$$\boldsymbol{\theta}_0 = \operatorname*{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} M_0(\boldsymbol{\theta}), \qquad M_0(\boldsymbol{\theta}) = \lim_{n \to \infty} M_n(\boldsymbol{\theta}), \qquad M_n(\boldsymbol{\theta}) = \mathbb{E}[m_n(\mathbf{z}, \boldsymbol{\theta})].$$

In other words, the setting considered in this section is one where $\hat{\boldsymbol{\theta}}_n$ approximately maximizes a function $\hat{M}_n$ whose population counterpart $M_n$ can be interpreted as a regularization (in the sense of Bickel and Li, 2006) of a function $M_0$ whose maximizer $\boldsymbol{\theta}_0$ is the object of interest. The additional flexibility (relative to the more traditional $M$-estimation setting of Section 2) afforded by the present setting is attractive because it allows us to formulate results that cover local $M$-estimators such as the conditional maximum score estimator of Honoré and Kyriazidou (2000). Studying this type of setting, Seo and Otsu (2018) gave conditions under which $\hat{\boldsymbol{\theta}}_n$ converges at a rate equal to the cube root of the "effective" sample size and has a limiting distribution of Chernoff (1964) type. Analogous conclusions will be drawn below, albeit under slightly different conditions.

For any $n$ and any $\delta > 0$, define $\mathcal{M}_n = \{m_n(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, $\bar{m}_n(\mathbf{z}) = \sup_{m \in \mathcal{M}_n} |m(\mathbf{z})|$, $\boldsymbol{\Theta}_0^\delta = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq \delta\}$, $\mathcal{D}_n^\delta = \{m_n(\cdot, \boldsymbol{\theta}) - m_n(\cdot, \boldsymbol{\theta}_0) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\delta\}$, and $\bar{d}_n^\delta(\mathbf{z}) = \sup_{d \in \mathcal{D}_n^\delta} |d(\mathbf{z})|$.

**Condition CRA (Cube Root Asymptotics)** For a positive $q_n$ with $r_n = \sqrt[3]{n q_n} \to \infty$, the following are satisfied:

(i) $\{\mathcal{M}_n : n \geq 1\}$ is uniformly manageable for the envelopes $\bar{m}_n$ and $q_n \mathbb{E}[\bar{m}_n(\mathbf{z})^2] = O(1)$. Also, $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})| = o(1)$ and, for every $\delta > 0$, $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta} \setminus \boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) < M_0(\boldsymbol{\theta}_0)$.

(ii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$ and, for some $\delta > 0$, $M_0$ and $M_n$ are twice continuously differentiable on $\boldsymbol{\Theta}_0^\delta$ and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\delta} \left\| \partial^2 [M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})] / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \right\| = o(1)$.

Also, $r_n||\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}|| = o(1)$ and $\mathbf{V}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(iii) For some $\delta > 0$, $\{\mathcal{D}_n^{\delta'} : n \geq 1, 0 < \delta' \leq \delta\}$ is uniformly manageable for the envelopes $\bar{d}_n^{\delta'}$ and $q_n \sup_{0<\delta'\leq\delta} \mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$.

(iv) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$, $q_n^2\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^3] + q_n^3 r_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = o(1)$, 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 $\mathbf{s} \neq \mathbf{t}$,

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

(v) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$,

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

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

To interpret Condition CRA, consider first the benchmark case where $m_n = m_0$ and $q_n = 1$. In this case, the condition is similar to, but slightly stronger than, assumptions (ii)-(vii) of the main theorem of Kim and Pollard (1990), to which the reader is referred for a definition of the term (uniformly) manageable. The most notable difference between Condition CRA and the assumptions employed by Kim and Pollard (1990) is probably that part (iv) contains assumptions about moments of orders three and four, that the displayed part of part (iv) is a locally uniform (with respect to $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$) version of its counterpart in Kim and Pollard (1990), and that (i) can be thought of as replacing the high level condition $\hat{\boldsymbol{\theta}}_n \to_{\mathbb{P}} \boldsymbol{\theta}_0$ of Kim and Pollard (1990) with more primitive conditions that imply it for approximate $M$-estimators. In all three cases, the purpose of strengthening the assumptions of Kim and Pollard (1990) is to be able to analyze the bootstrap.

More generally, Condition CRA can be interpreted as an $n$-varying version of a suitably (for the purpose of analyzing the bootstrap) strengthened version of the assumptions of Kim and Pollard (1990). The differences between Condition CRA and the $i.i.d.$ version of the conditions in Seo and Otsu (2018) seem mostly technical in nature, but for completeness we highlight two differences here. First, to handle dependent data Seo and Otsu (2018) control the complexity of various function classes using the concept of bracketing entropy. In contrast, because we assume random sampling we can follow Kim and Pollard (1990) and obtain maximal inequalities using bounds on uniform entropy numbers implied by the concept of (uniform) manageability. Second, whereas Seo and Otsu (2018) controls the bias of $\hat{\boldsymbol{\theta}}_n$ through a locally uniform bound on $M_n - M_0$, Condition CRA controls the bias the first and second derivatives of $M_n - M_0$.

Under Condition CRA, the effective sample size is given by $nq_n$. In perfect agreement with Seo and Otsu (2018), it turns out that if $\hat{\boldsymbol{\theta}}_n$ is an approximate maximizer of $\hat{M}_n$, then

$$r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\}, \tag{7}$$

where $\mathcal{Q}_0(\mathbf{s}) = -\mathbf{s}'\mathbf{V}_0\mathbf{s}/2$, and where $\mathcal{G}_0$ is a zero-mean Gaussian process with $\mathcal{G}_0(0) = 0$ and covariance kernel $\mathcal{C}_0$. The heuristics of the previous section are rate-adaptive, so once again it stands to reason that if $\tilde{V}_n$ is a consistent estimator of $V_0$, then a consistent distributional approximation can be based on an approximate maximizer $\tilde{\boldsymbol{\theta}}_n^*$ of

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

where $\mathbf{z}_{1,n}^*, \dots, \mathbf{z}_{n,n}^*$ is a random sample from the empirical distribution of $\mathbf{z}_1, \dots \mathbf{z}_n$.

Following van der Vaart (1998, Chapter 23), we say that our bootstrap-based estimator of the distribution of $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is consistent if

$$\sup_{\mathbf{t}\in\mathbb{R}^d}\left|\mathbb{P}^*[r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \leq \mathbf{t}] - \mathbb{P}[r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leq \mathbf{t}]\right| \to_{\mathbb{P}} 0. \tag{8}$$

Because the limiting distribution in (7) is continuous, this consistency property implies consistency of bootstrap-based confidence intervals (e.g., van der Vaart, 1998, Lemma 23.3). Moreover, continuity of the limiting distribution implies that (8) holds provided the estimator $\tilde{\boldsymbol{\theta}}_n^*$ satisfies the following bootstrap counterpart of (7):

$$r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \rightsquigarrow_{\mathbb{P}} \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\}.$$

Theorem 1, our main result, gives sufficient conditions for this to occur.

**Theorem 1** *Suppose Condition CRA holds. If* $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$ *and if*

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

*then (8) holds.*

To implement the bootstrap-based approximation to the distribution of $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, only a consistent estimator of $\mathbf{V}_0$ is needed. A generic estimator based on numerical derivatives is $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$, the matrix whose element $(k, l)$ is given by

$$\begin{aligned}\tilde{V}_{n,kl}^{\mathtt{ND}} \quad = \quad -\frac{1}{4\epsilon_n^2}\Big[&\hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{e}_k\epsilon_n + \mathbf{e}_l\epsilon_n) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{e}_k\epsilon_n - \mathbf{e}_l\epsilon_n) \\ &- \hat{M}_n(\hat{\boldsymbol{\theta}}_n - \mathbf{e}_k\epsilon_n + \mathbf{e}_l\epsilon_n) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n - \mathbf{e}_k\epsilon_n - \mathbf{e}_l\epsilon_n)\Big],\end{aligned}$$

where $\mathbf{e}_k$ is the $k$th unit vector in $\mathbb{R}^d$ and where $\epsilon_n$ is a positive tuning parameter. Conditions under which this estimator is consistent are given in the following lemma.

**Lemma 1** *Suppose Condition CRA holds and that* $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$. *If* $\epsilon_n \to 0$ *and if* $r_n\epsilon_n \to \infty$, *then* $\tilde{\mathbf{V}}_n^{\mathtt{ND}} \to_{\mathbb{P}} \mathbf{V}_0$.

Plausibility of the high-level condition $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ follows from (7). More generally, if only consistency is assumed on the part of $\hat{\boldsymbol{\theta}}_n$, then $\tilde{\mathbf{V}}_n^{\mathtt{ND}} \to_{\mathbb{P}} \mathbf{V}_0$ holds provided $\epsilon_n \to 0$ and $||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0||/\epsilon_n \to_{\mathbb{P}} 0$. The proof of the lemma goes beyond consistency and develops a Nagar-type mean squared error (MSE) expansion for $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ under the additional assumption that $M_0$ is four times continuously differentiable near $\boldsymbol{\theta}_0$. The approximate MSE can be minimized by choosing $\epsilon_n$ proportional to $r_n^{-3/7}$, the optimal factor of proportionality being a functional of the covariance kernel $\mathcal{C}_0$ and the fourth order derivatives of $M_0$ evaluated at $\boldsymbol{\theta}_0$. For details, see the supplemental appendix (Section A.3, Theorem A.3), which also contains a brief discussion of alternative generic estimators of $\mathbf{V}_0$.

We close this section by summarizing the algorithm for our proposed bootstrap-based distributional approximation.

**Bootstrap-Based Approximation** Let the notation and conditions in Theorem 1 hold.

*Step 1.* Compute $\hat{\boldsymbol{\theta}}_n$ and $\tilde{\mathbf{V}}_n$ using the original sample $\mathbf{z}_1, \ldots, \mathbf{z}_n$.

*Step 2.* Compute $\tilde{M}_n^*(\boldsymbol{\theta})$ and let $\tilde{\boldsymbol{\theta}}_n^*$ be an approximate maximizer thereof, both constructed using the nonparametric bootstrap sample $\mathbf{z}_{1,n}^*, \ldots, \mathbf{z}_{n,n}^*$. (Note that $\tilde{\mathbf{V}}_n$ is not recomputed at this step.)

*Step 3.* Repeat Step 2 $B$ times, and then compute the empirical distribution of $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)$.

# 4 Examples

This section briefly discusses two econometric examples covered by our main result, namely the maximum score estimator of Manski (1975) and the conditional maximum score estimator of Honoré and Kyriazidou (2000). From the perspective of this paper, the main difference between these examples is that (only) the latter corresponds to a situation where $m_n$ depends on $n$.

## 4.1 The Maximum Score Estimator

To describe a version of the maximum score estimator of Manski (1975), suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a random sample of $\mathbf{z} = (y, \mathbf{x}')'$ generated by the binary response model

$$y = \mathbb{1}(\boldsymbol{\beta}_0'\mathbf{x} + u \geq 0), \qquad F_{u|\mathbf{x}}(0|\mathbf{x}) = 1/2, \tag{9}$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^{d+1}$ is an unknown parameter of interest, $\mathbf{x} \in \mathbb{R}^{d+1}$ is a vector of covariates, and $F_{u|\mathbf{x}}(\cdot|\mathbf{x})$ is the conditional cumulative distribution function of the unobserved error term $u$ given $\mathbf{x}$. Following Abrevaya and Huang (2005), we employ the parameterization $\boldsymbol{\beta}_0 = (1, \boldsymbol{\theta}_0')'$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is unknown. In other words, we assume that the first element of $\boldsymbol{\beta}_0$ is positive and then normalize the (unidentified) scale of $\boldsymbol{\beta}_0$ by setting its first element equal to unity. Partitioning $\mathbf{x}$ conformably with $\boldsymbol{\beta}_0$ as $\mathbf{x} = (x_1, \mathbf{x}_2')'$, a maximum score estimator of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ is any $\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}$ approximately maximizing $\hat{M}_n$ for

$$m_n(\mathbf{z}, \boldsymbol{\theta}) = m^{\mathtt{MS}}(\mathbf{z}, \boldsymbol{\theta}) = (2y - 1)\mathbb{1}(x_1 + \boldsymbol{\theta}'\mathbf{x}_2 \geq 0).$$

Regarded as a member of the class of $M$-estimators exhibiting cube root asymptotics, the maximum score estimator is representative in a couple of respects. First, under easy-to-interpret primitive conditions the estimator is covered by the results of Section 3. In particular, because

$m_n$ does not depend on $n$ we can set $q_n = 1$ when formulating primitive conditions for Condition CRA; for details, see the supplemental appendix (Section A.4). Second, in addition to the generic estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ discussed above, the maximum score estimator admits an example-specific consistent estimator of the $\mathbf{V}_0$ associated with it. Let

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}} = -\frac{1}{n}\sum_{i=1}^{n}(2y_i - 1)\dot{K}_n(x_{1i} + \boldsymbol{\theta}'\mathbf{x}_{2i})\mathbf{x}_{2i}\mathbf{x}_{2i}'\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}},$$

where, for a kernel function $K$ and a bandwidth $h_n$, $\dot{K}_n(u) = d\dot{K}_n(u)/du$ and $K_n(u) = K(u/h_n)/h_n$. As defined, $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ is simply minus the second derivative, evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}$, of the criterion function associated with the smoothed maximum score estimator of Horowitz (1992). The estimator $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ is consistent under mild conditions on $h_n$ and $K$; for details, see the supplemental appendix (Section A.4, Lemma MS), which also reports the results of a Monte Carlo experiment evaluating the performance of our proposed inference procedure.

### 4.2 The Conditional Maximum Score Estimator

Consider the dynamic binary response model

$$Y_t = \mathbb{1}(\boldsymbol{\beta}_0'\mathbf{X}_t + \gamma_0 Y_{t-1} + \alpha + u_t \geq 0), \qquad t = 1, 2, 3,$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ and $\gamma_0 \in \mathbb{R}$ are unknown parameters of interest, $\alpha$ is an unobserved (time-invariant) individual-specific effect, and $u_t$ is an unobserved error term. Honoré and Kyriazidou (2000) analyzed this model and gave conditions under which $\boldsymbol{\beta}_0$ and $\gamma_0$ are identified up to scale. Assuming these conditions hold and assuming the first element of $\boldsymbol{\beta}_0$ is positive, we can normalize that element to unity and employ the parameterization $(\boldsymbol{\beta}_0', \gamma_0)' = (1, \boldsymbol{\theta}_0')'$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is unknown.

To describe a version of the conditional maximum score estimator of Honoré and Kyriazidou (2000), partition $\mathbf{X}_t$ after the first element as $\mathbf{X}_t = (X_{1t}, \mathbf{X}_{2t}')'$ and define $\mathbf{z} = (y, x_1, \mathbf{x}_2', \mathbf{w}')'$, where $y = y_2 - y_1$, $x_1 = X_{12} - X_{11}$, $\mathbf{x}_2 = ((\mathbf{X}_{22} - \mathbf{X}_{21})', y_3 - y_0)'$, and $\mathbf{w} = \mathbf{X}_2 - \mathbf{X}_3$. Assuming $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a random sample of $\mathbf{z}$, a conditional maximum score estimator of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ is any

$\hat{\boldsymbol{\theta}}_n^{\texttt{CMS}}$ approximately maximizing $\hat{M}_n$ for

$$m_n(\mathbf{z},\boldsymbol{\theta}) = m_n^{\texttt{CMS}}(\mathbf{z},\boldsymbol{\theta}) = y\mathbb{1}(x_1 + \boldsymbol{\theta}'\mathbf{x}_2 \geq 0)L_n(\mathbf{w}),$$

where, for a kernel function $L$ and a bandwidth $b_n$, $L_n(\mathbf{w}) = L(\mathbf{w}/b_n)/b_n^d$.

Through its dependence on $b_n$, the function $m_n^{\texttt{CMS}}$ depends on $n$ in a non-negligible way. In particular, the effective sample size is $nb_n^d$ (rather than $n$) in the current setting, so to the extent that they exist one would expect primitive sufficient conditions for Condition CRA to include $q_n = b_n^d$ in this example. Apart from this predictable change, the properties of the conditional maximum score estimator $\hat{\boldsymbol{\theta}}_n^{\texttt{CMS}}$ turn out to be qualitatively similar to those of $\hat{\boldsymbol{\theta}}_n^{\texttt{MS}}$. To be specific, under regularity conditions the conditional maximum score estimator is covered by the results of Section 3 and an example-specific alternative (of smoothed maximum score type) to the generic numerical derivative estimator $\tilde{\mathbf{V}}_n^{\texttt{ND}}$ is available; for details, see the supplemental appendix (Section A.5).

# 5 Discussion

The applicability of the procedure proposed in this note extends beyond the estimators covered by Theorem 1. For instance, it is not hard to show that our bootstrap-based distributional approximation is consistent also in the more standard case where $m_n(\mathbf{z},\boldsymbol{\theta})$ is sufficiently smooth in $\boldsymbol{\theta}$ to ensure that an approximate maximizer of $\hat{M}_n$ is asymptotically normal and that the nonparametric bootstrap is consistent. In fact, $\tilde{\boldsymbol{\theta}}_n^*$ is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_n^*$ in that standard case, so our procedure can be interpreted as a modification of the nonparametric bootstrap that is designed to be "robust" to the types of non-smoothness that give rise to cube root asymptotics.

Moreover, and perhaps more importantly, the idea of reshaping can be used to achieve consistency of bootstrap-based approximations to the distributions of certain estimators which are not of $M$-estimator type, yet exhibit cube root asymptotics and have the feature that the standard bootstrap-based approximations to their distribution are known to be inconsistent. In particular, the supplemental appendix (Section A.6) shows how the idea of reshaping a process can be used to achieve consistency on the part of a bootstrap-based approximation to the distribution of the celebrated ($\sqrt[3]{n}$-consistent) isotonic density estimator of Grenander (1956).[1]

This note is not the first to propose a consistent resampling-based distributional approximation for cube root consistent estimators. For cube root asymptotic problems, the best known consistent alternative to the nonparametric bootstrap is probably subsampling (Politis and Romano, 1994), whose applicability was pointed out by Delgado, Rodriguez-Poo, and Wolf (2001). A related method is the rescaled bootstrap (Dümbgen, 1993), whose validity in cube root asymptotic $M$-estimation problems was established recently by Hong and Li (2017). In addition, case-specific smooth bootstrap methods have been proposed for leading examples such as maximum score estimation (Patra, Seijo, and Sen, 2015) and isotonic density estimation (Kosorok, 2008; Sen, Banerjee, and Woodroofe, 2010). Like ours, each of these methods can be viewed as offering a "robust" alternative to the nonparametric bootstrap but, unlike ours, they all achieve consistency by modifying the distribution used to generate the bootstrap counterpart of the estimator whose distribution is being approximated. In contrast, our method achieves consistency by means of an analytic modification to the objective function used to construct the bootstrap-based distributional approximation.

As pointed out by two referees and the coeditor, an alternative interpretation of our approach is available. Restating the result in (7) as

$$r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \mathcal{S}_0(\mathcal{G}_0), \qquad \mathcal{S}_0(\mathcal{G}) = \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}(\mathbf{s})\},$$

our procedure approximates the distribution of $\mathcal{S}_0(\mathcal{G}_0)$ by that of $\tilde{\mathcal{S}}_n(\hat{G}_n^*)$, where

$$\hat{G}_n^*(\mathbf{s}) = r_n^2[\hat{M}_n^*(\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1}) - \hat{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1}) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n)]$$

is a bootstrap process whose distribution approximates that of $\mathcal{G}_0(\mathbf{s})$ and where

$$\tilde{\mathcal{S}}_n(\mathcal{G}) = \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\tilde{Q}_n(\mathbf{s}) + \mathcal{G}(\mathbf{s})\}, \qquad \tilde{Q}_n(\mathbf{s}) = -\frac{1}{2}\mathbf{s}'\tilde{\mathbf{V}}_n\mathbf{s},$$

is an estimator of $\mathcal{S}_0(\mathcal{G})$. In other words, our procedure replaces the functional $\mathcal{S}_0$ with a consistent

[1] The asymptotic properties of the Grenander estimator have been studied by Prakasa Rao (1969), Groeneboom (1985), and Kim and Pollard (1990), among others. Inconsistency of the standard bootstrap-based approximation to the distribution of the Grenander estimator has been pointed out by Kosorok (2008) and Sen, Banerjee, and Woodroofe (2010), among others.

estimator (namely, $\tilde{\mathcal{S}}_n$) and its random argument $\mathcal{G}_0$ with a bootstrap approximation (namely, $\hat{G}_n^*$). Similar constructions have been successfully applied in other settings, two relatively recent examples being Andrews and Soares (2010) and Fang and Santos (2016).

Finally, Seo and Otsu (2018) give conditions under which results of the form (7) can be obtained also when the data is weakly dependent. It seems plausible that a version of our procedure, implemented with resampling procedure suitable for dependent data, can be shown to be consistent under similar conditions, but it is beyond the scope of this note to substantiate that conjecture.

## References


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

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Bickel, P. J., and B. Li (2006): "Regularization in Statistics," *Test*, 15(2), 271–344.

Chernoff, H. (1964): "Estimation of the Mode," *Annals of the Institute of Statistical Mathematics*, 16(1), 31–41.

Delgado, M. A., J. M. Rodriguez-Poo, and M. Wolf (2001): "Subsampling Inference in Cube Root Asymptotics with an Application to Manski's Maximum Score Estimator," *Economics Letters*, 73(2), 241–250.

Dümbgen, L. (1993): "On Nondifferentiable Functions and the Bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Fang, Z., and A. Santos (2016): "Inference on Directionally Differentiable Functions," arXiv:1404.3763.

Grenander, U. (1956): "On the Theory of Mortality Measurement: Part II," *Scandinavian Actuarial Journal*, 39(2), 125–153.

Groeneboom, P. (1985): "Estimating a Monotone Density," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, pp. 539–555. Institute of Mathematical Statistics.

Hong, H., and J. Li (2017): "The Numerical Delta Method and Bootstrap," Working Paper, Stanford University.

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

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60(3), 505–531.

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

Kosorok, M. R. (2008): "Bootstrapping the Grenander Estimator," in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pp. 282–292. Institute of Mathematical Statistics.

Léger, C., and B. MacGibbon (2006): "On the Bootstrap in Cube Root Asymptotics," *Canadian Journal of Statistics*, 34(1), 29–44.

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of econometrics*, 3(3), 205–228.

Patra, R. K., E. Seijo, and B. Sen (2015): "A Consistent Bootstrap Procedure for the Maximum Score Estimator," arXiv:1105.1976.

Politis, D. N., and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22(4), 2031–2050.

Prakasa Rao, B. (1969): "Estimation of a Unimodal Density," *Sankhya: The Indian Journal of Statistics, Series A*, 31(1), 23–36.

Sen, B., M. Banerjee, and M. Woodroofe (2010): "Inconsistency of Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38(4), 1953–1977.

Seo, M. H., and T. Otsu (2018): "Local M-Estimation with Discontinuous Criterion for Dependent and Limited Observations," *Annals of Statistics*, 46(1), 344–369.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer.

# Supplemental to "Bootstrap-Based Inference for Cube Root Consistent Estimators"*

Matias D. Cattaneo[†] Michael Jansson[‡] Kenichi Nagasawa[§]

May 30, 2018

**Abstract**

This supplemental appendix contains proofs of all the results stated in the paper, as well as other theoretical results that may be of independent interest. It also offers more details on the examples discussed in the paper. Finally, details on implementation issues are given, including mean square error expansions of drift estimators, associated MSE-optimal tuning parameter choices, and rule-of-thumb implementations thereof.

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

[†]Department of Economics and Department of Statistics, University of Michigan.

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

[§]Department of Economics, University of Michigan.

## Contents

## A.1 Setup

Suppose the estimator $\hat{\boldsymbol{\theta}}_n$ is an approximate maximizer (with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$) of

$$\hat{M}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} m_n(\mathbf{z}_i, \boldsymbol{\theta}),$$

where $m_n$ is a known function and where $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a random sample of a random vector $\mathbf{z}$. Also, suppose the estimand $\boldsymbol{\theta}_0$ admits the characterization

$$\boldsymbol{\theta}_0 = \operatorname*{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} M_0(\boldsymbol{\theta}), \qquad M_0(\boldsymbol{\theta}) = \lim_{n \to \infty} M_n(\boldsymbol{\theta}), \qquad M_n(\boldsymbol{\theta}) = \mathbb{E}[m_n(\mathbf{z}, \boldsymbol{\theta})].$$

Finally, suppose $\tilde{\boldsymbol{\theta}}_n^*$ is an approximate maximizer of

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

where $\mathbf{z}_{1,n}^*, \ldots, \mathbf{z}_{n,n}^*$ is a random sample from the empirical distribution of $\mathbf{z}_1, \ldots \mathbf{z}_n$ and where $\tilde{\mathbf{V}}_n$ is a consistent estimator of $\mathbf{V}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$.

Our main result gives conditions under which the distribution of $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ is well approximated by the bootstrap distribution of $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n$. More precisely, we give conditions under which the following algorithm produces a consistent estimator of the distribution of $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, where $r_n$ is defined below:

*Step 1.* Given the sample $\mathbf{z}_1, \ldots, \mathbf{z}_n$, calculate $\hat{\boldsymbol{\theta}}_n$, $\tilde{\mathbf{V}}_n$, and $\hat{F}_n(\cdot) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(z_i \leq \cdot)$.

*Step 2.* Draw a bootstrap sample $\mathbf{z}_{1,n}^*, \ldots, \mathbf{z}_{n,n}^*$ from $\hat{F}_n$.

*Step 3.* Compute $\tilde{M}_n^*(\boldsymbol{\theta})$ and let $\tilde{\boldsymbol{\theta}}_n^*$ be an approximate maximizer thereof.

*Step 4.* Repeat the steps 2 and 3 $B$ times and compute the empirical distribution of $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)$.

To implement the procedure, an estimator $\tilde{\mathbf{V}}_n$ of $\mathbf{V}_0$ is needed. The specific estimator described in the paper is the numerical derivative-based estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$, the matrix whose element $(k, l)$ is given by

$$\begin{aligned} \tilde{V}_{n,kl}^{\mathtt{ND}} \quad = \quad -\frac{1}{4\epsilon_n^2} \Big[ & \hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{e}_k\epsilon_n + \mathbf{e}_l\epsilon_n) - \hat{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{e}_k\epsilon_n - \mathbf{e}_l\epsilon_n) \\ & -\hat{M}_n(\hat{\boldsymbol{\theta}}_n - \mathbf{e}_k\epsilon_n + \mathbf{e}_l\epsilon_n) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n - \mathbf{e}_k\epsilon_n - \mathbf{e}_l\epsilon_n) \Big], \end{aligned}$$

where $\mathbf{e}_k$ is the $k$th unit vector in $\mathbb{R}^d$ and where $\epsilon_n$ is a positive tuning parameter.

## A.2 Assumptions

Our main assumptions are collected in Condition CRA. To state the condition, for any $n$ and any $\delta > 0$, define $\mathcal{M}_n = \{m_n(\cdot, \boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, $\bar{m}_n(\mathbf{z}) = \sup_{m \in \mathcal{M}_n} |m(\mathbf{z})|$, $\boldsymbol{\Theta}_0^\delta = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq \delta\}$,

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

**Condition CRA (Cube Root Asymptotics)** For a positive $q_n$ with $r_n = \sqrt[3]{nq_n} \to \infty$, the following are satisfied:

(i) $\{\mathcal{M}_n : n \geq 1\}$ is uniformly manageable for the envelopes $\bar{m}_n$ and $q_n\mathbb{E}[\bar{m}_n(\mathbf{z})^2] = O(1)$. Also, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})| = o(1)$ and, for every $\delta > 0$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) < M_0(\boldsymbol{\theta}_0)$.

(ii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$ and, for some $\delta > 0$, $M_0$ and $M_n$ are twice continuously differentiable on $\boldsymbol{\Theta}_0^\delta$ and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta} \left\|\partial^2[M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right\| = o(1)$.
Also, $\sqrt[3]{nq_n}||\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}|| = o(1)$ and $\mathbf{V}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(iii) For some $\delta > 0$, $\{\mathcal{D}_n^{\delta'} : n \geq 1, 0 < \delta' \leq \delta\}$ is uniformly manageable for the envelopes $\bar{d}_n^{\delta'}$ and $q_n \sup_{0<\delta'\leq\delta} \mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$.

(iv) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$, $q_n^2\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^3] + q_n^3 r_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = o(1)$, 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 $\mathbf{s} \neq \mathbf{t}$,

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

(v) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$,

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

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

Suppose $m_n = m_0$. In that case, we can set $q_n = 1$ and simplify Condition CRA somewhat. For any $n$ and any $\delta > 0$, define $\mathcal{M}_0 = \{m_0(\cdot,\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, $\bar{m}_0(\mathbf{z}) = \sup_{m\in\mathcal{M}_0} |m(\mathbf{z})|$, $\mathcal{D}_0^\delta = \{m_0(\cdot,\boldsymbol{\theta}) - m_0(\cdot,\boldsymbol{\theta}_0) : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0^\delta\}$, and $\bar{d}_0^\delta(\mathbf{z}) = \sup_{d\in\mathcal{D}_0^\delta} |d(\mathbf{z})|$.

**Condition CRA$_0$ (Cube Root Asymptotics, benchmark case)** For $r_n = \sqrt[3]{n} \to \infty$, 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 $\delta > 0$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) < M_0(\boldsymbol{\theta}_0)$.

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

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

(iv) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$, $\mathbb{E}[\bar{d}_0^{\delta_n}(\mathbf{z})^3] + r_n^{-1}\mathbb{E}[\bar{d}_0^{\delta_n}(\mathbf{z})^4] = o(1)$, 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 $\mathbf{s} \neq \mathbf{t}$,

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

(v) For every positive $\delta_n$ with $\delta_n = O(r_n^{-1})$,

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

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

## A.3 Main Results

**Theorem A.1** *Suppose Condition CRA holds. If*

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

*then*

$$r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\},$$

*where* $\mathcal{Q}_0(\mathbf{s}) = -\mathbf{s}'\mathbf{V}_0\mathbf{s}/2$, *and where* $\mathcal{G}_0$ *is a zero-mean Gaussian process with* $\mathcal{G}_0(0) = 0$ *and covariance kernel* $\mathcal{C}_0$.

**Theorem A.2** *Suppose Condition CRA holds. If* $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$, $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$, *and if*

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

*then*

$$r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \rightsquigarrow_{\mathbb{P}} \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{Q}_0(\mathbf{s}) + \mathcal{G}_0(\mathbf{s})\},$$

*where* $\mathcal{Q}_0(\mathbf{s})$ *and* $\mathcal{G}_0$ *are as in Theorem A.1.*

Combining Theorems A.1 and A.2, we obtain the main result of the paper.

**Corollary A.1** *Suppose Condition CRA holds. If* $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$ *and if*

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

*then*

$$\sup_{\mathbf{t}\in\mathbb{R}^d} \left| \mathbb{P}^*[r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \leq \mathbf{t}] - \mathbb{P}[r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leq \mathbf{t}] \right| \to_{\mathbb{P}} 0.$$

The following result gives conditions under which $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ is consistent and furthermore presents results that can be used to obtain a data-based selector of the tuning parameter $\epsilon_n$. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_d)'$.

**Theorem A.3** *Suppose Condition CRA holds and that* $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$. *If* $\epsilon_n \to 0$ *and if* $r_n\epsilon_n \to \infty$, *then* $\tilde{\mathbf{V}}_n^{\mathtt{ND}} \to_{\mathbb{P}} \mathbf{V}_0$.

*If, in addition, for some* $\delta > 0$, $M_0$ *and* $M_n$ *are four times continuously differentiable on* $\boldsymbol{\Theta}_0^\delta$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta} |\partial^4[M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})]/\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}| = o(1)$ *for all* $j_1, j_2, j_3, j_4 \in \{1, 2, \cdots, d\}$, *and* $\mathcal{C}_0(\mathbf{s}, -\mathbf{s}) = 0$ *and* $\mathcal{C}_0(\mathbf{s}, \mathbf{t}) = \mathcal{C}_0(-\mathbf{s}, -\mathbf{t})$ *for all* $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, *then* $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ *admits an approximation* $\check{\mathbf{V}}_n^{\mathtt{ND}}$ *satisfying*

$$\tilde{\mathbf{V}}_n^{\mathtt{ND}} - \check{\mathbf{V}}_n^{\mathtt{ND}} = O_{\mathbb{P}}\left(r_n^{-2}\epsilon_n^{-2}\right) + o(\epsilon_n^2) + O(r_n^{-1})$$

*and*

$$\mathbb{E}[||\check{\mathbf{V}}_n^{\mathtt{ND}} - \mathbf{V}_n||^2] = \epsilon_n^4(\sum_{k=1}^d\sum_{l=1}^d \mathsf{B}_{kl}^2) + r_n^{-3}\epsilon_n^{-3}(\sum_{k=1}^d\sum_{l=1}^d \mathsf{V}_{kl}) + o(\epsilon_n^2 + r_n^{-3}\epsilon_n^{-3}),$$

*where* $\mathbf{V}_n = -\partial^2 M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ *and where*

$$\begin{aligned}
\mathsf{B}_{kl} &= \frac{1}{6}\left(\frac{\partial^4}{\partial\theta_k^3\partial\theta_l}M_0(\boldsymbol{\theta}_0) + \frac{\partial^4}{\partial\theta_k\partial\theta_l^3}M_0(\boldsymbol{\theta}_0)\right),\\
\mathsf{V}_{kl} &= \frac{1}{8}[\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, \mathbf{e}_k + \mathbf{e}_l) + \mathcal{C}_0(\mathbf{e}_k - \mathbf{e}_l, \mathbf{e}_k - \mathbf{e}_l)\\
&\qquad - 2\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, \mathbf{e}_k - \mathbf{e}_l) - 2\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, -\mathbf{e}_k + \mathbf{e}_l)].
\end{aligned}$$

Using this theorem, we immediately obtain the asymptotic MSE-optimal tuning parameter choice

$$\epsilon_n^{\mathtt{AMSE}} = \left(\frac{3\sum_{k=1}^d\sum_{l=1}^d \mathsf{V}_{kl}}{4\sum_{k=1}^d\sum_{l=1}^d \mathsf{B}_{kl}^2}\right)^{1/7} r_n^{-3/7}.$$

As the notation suggests, the constants $\mathsf{B}_{kl}$ and $\mathsf{V}_{kl}$ correspond to element $(k, l)$ of $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$, so the asymptotic MSE-optimal tuning parameter choice for $\tilde{V}_{n,kl}^{\mathtt{ND}}$ is

$$\epsilon_{n,kl}^{\mathtt{AMSE}} = \left(\frac{3\mathsf{V}_{kl}}{4\mathsf{B}_{kl}^2}\right)^{1/7} r_n^{-3/7}.$$

**Remark** There are other possible ways of constructing generic and automatic consistent estimators of $\mathbf{V}(\boldsymbol{\theta})$. Here we briefly mention two generic alternatives, but we do not study their formal statistical properties to conserve space.

1. *Jittering Smoothing.* Define $m_{n,\eta}(\mathbf{z}_i, \boldsymbol{\theta}) = \int m_n(\mathbf{z}_i, \boldsymbol{\theta} + \epsilon_n\eta)dF_\eta(\eta)$, where $\eta$ is a random variable with absolutely continuous distribution function $F_\eta(\eta)$ satisfying $\int \eta dF_\eta(\eta) = 0$ and $\int \eta^2 dF_\eta(\eta) = 1$. By choosing $F_\eta(\eta)$ appropriately, $\boldsymbol{\theta} \mapsto m_{n,\eta}(\mathbf{z}_i, \boldsymbol{\theta})$ can be taken to be two-times continuously differentiable. Therefore, a consistent plug-in estimator of $\mathbf{V}(\boldsymbol{\theta})$ can easily be constructed by differentiating $m_{n,\eta}(\mathbf{z}_i, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, under appropriate conditions on $\epsilon_n \to 0$.

2. *Local Polynomial Smoothing.* An alternative is to smooth-out the objective function directly. Suppose $\{\boldsymbol{\theta}_\ell : 1 \leq \ell \leq L\}$ is a grid of points near $\boldsymbol{\theta}$ on $\boldsymbol{\Theta}$, with the property that $\boldsymbol{\theta}_\ell - \boldsymbol{\theta}_{\ell-1} = o(1)$, then a local polynomial smoothing approach is

$$\hat{\boldsymbol{\gamma}} = \operatorname*{argmax}_{\boldsymbol{\gamma}\in\mathbb{R}^p} \sum_{\ell=1}^{L} (\hat{M}_n(\boldsymbol{\theta}) - \mathbf{r}_p(\boldsymbol{\theta}_\ell - \boldsymbol{\theta})'\boldsymbol{\gamma})^2 K\left(\frac{\boldsymbol{\theta}_\ell - \boldsymbol{\theta}}{\epsilon_n}\right)$$

where $\mathbf{r}_p(\boldsymbol{\theta})$ represents a full polynomial expansion of order $p \geq 2$, and $\epsilon_n$ is a positive vanishing bandwidth sequence. By choosing the appropriate elements of $\hat{\boldsymbol{\gamma}}$, we obtain an alternative estimator of $\mathbf{V}(\boldsymbol{\theta})$.

## A.4 Example: Maximum Score Estimation

Our results apply directly to Manski's Maximum Score Estimator (Manski, 1975, 1985). The econometric model is

$$y = \mathbb{1}(\mathbf{x}'\boldsymbol{\beta}_0 + u \geq 0), \qquad F_{u|\mathbf{x}}(0|\mathbf{x}) = 1/2,$$

where $\mathbf{z} = (y, \mathbf{x}')'$ and $\mathbf{x} = (x_1, \mathbf{x}_2')'$ is a $(d+1)$-dimensional vector of covariates. For identifiability, we need a normalization of $\boldsymbol{\beta}_0$. As in Abrevaya and Huang (2005), we employ $\boldsymbol{\beta}_0 = (1, \boldsymbol{\theta}_0')'$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is unknown.

Then, a maximum score estimator of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^d$ is any $\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}$ approximately maximizing $\hat{M}_n$ for

$$m_n(\mathbf{z}, \boldsymbol{\theta}) = m^{\mathtt{MS}}(\mathbf{z}, \boldsymbol{\theta}) = (2y - 1)\mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0).$$

Therefore, under random sampling, the maximum score estimator is computed as the approximate $M$-estimator

$$\hat{\boldsymbol{\beta}}_n \approx \operatorname*{argmax}_{\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{n}\sum_{i=1}^{n}(2y_i - 1)\mathbb{1}(\mathbf{x}_i'\boldsymbol{\beta} \geq 0),$$

where $\mathcal{B}$ is the parameter space such that $\mathcal{B} \subset \{\boldsymbol{\beta} \in \mathbb{R}^{d+1} : |\mathbf{e}_1'\boldsymbol{\beta}| = 1\}$.

### A.4.1 Bootstrap-Based Inference

To verify the high-level conditions in Condition $\text{CRA}_0$, we assume the following.

**Condition MS** Suppose the maximum score model is as above.

(i) $\text{Median}(u|\mathbf{x}) = 0$ and $0 < \mathbb{P}(y = 1|\mathbf{x}) < 1$ almost surely. The conditional distribution function of $u$ given $\mathbf{x}$, denoted by $F_{u|x_1,\mathbf{x}_2}(u|x_1, \mathbf{x}_2)$, is bounded and $S_F$ times continuously differentiable in $u$ and $x_1$ with bounded derivatives for some $S_F \geq 1$.

(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.

(iii) The set $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \in \mathbb{R}^d : (1, \boldsymbol{\theta}')' \in \mathcal{B}\}$ is compact and $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.

(iv) $M_0^{\mathtt{MS}}(\boldsymbol{\theta}) = \mathbb{E}[(2y - 1)\mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0)]$ is $S$ times continuously differentiable near $\boldsymbol{\theta}_0$ for some $S \geq 2$ and $\mathbf{V}_0^{\mathtt{MS}} = -\partial^2 M_0^{\mathtt{MS}}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(v) The conditional density of $x_1$ given $\mathbf{x}_2$, denoted by $f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)$, is bounded and $S_f$ times continuously differentiable in $x_1$ with bounded derivatives for some $S_f \geq 1$.

Under these assumptions, we can deduce the following result.

**Corollary MS** *If Condition MS holds, then*

$$\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}} - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d} \left\{ -\frac{1}{2}\mathbf{s}'\mathbf{V}_0^{\mathtt{MS}}\mathbf{s} + \mathcal{G}_0(\mathbf{s}) \right\}$$

*where $\mathcal{G}_0$ is a mean-zero Gaussian process indexed by $\mathbb{R}^d$ with covariance kernel $\mathcal{C}_0^{\mathtt{MS}}(\cdot,\cdot)$, and*

$$\mathbf{V}_0^{\mathtt{MS}} = 2\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2' f_{z|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) f_{\varepsilon|z,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\right],$$

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

*Furthermore, if a consistent estimator for the Hessian term is available (i.e., $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0^{\mathtt{MS}}$ for some estimator $\tilde{\mathbf{V}}_n$), then the reshaped bootstrap consistently estimates the large sample distribution of $\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}} - \boldsymbol{\theta}_0)$.*

### A.4.2 Drift Estimation

Here we discuss an example-specific construction of $\tilde{\mathbf{V}}_n$ for the maximum score estimator and sufficient conditions for its consistency. Our proposed estimator for maximum score estimation is

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}} = \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}), \qquad \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}(2y_i - 1)\dot{K}((x_{1i} + \mathbf{x}_{2i}'\boldsymbol{\theta})/h_n)\mathbf{x}_{2i}\mathbf{x}_{2i}'h_n^{-2} \tag{A.1}$$

where $h_n = o(1)$ is a bandwidth, $K$ is a kernel function, and $\dot{K}(u) = dK(u)/du$.

To analyze the properties of this estimator, we impose the following conditions on the kernel function.

**Condition K**

(i) $\int_{\mathbb{R}} K(v)dv = 1$ and $\lim_{|v|\to\infty} vK(v) = 0$.

(ii) For all $v_1, v_2 \in \mathbb{R}$, $|\dot{K}(v_1) - \dot{K}(v_2)| \leq B(v_1)|v_1 - v_2|$ with $\int_{\mathbb{R}} B(v)dv + \int_{\mathbb{R}} B(v)^2 dv < \infty$.

(iii) $\int_{\mathbb{R}} |\dot{K}(v)|^2 dv + \int_{\mathbb{R}} |v\dot{K}(v)|dv < \infty$.

(iv) $\lim_{|v|\to\infty} v^2 K(v) = 0$, $\int_{\mathbb{R}} vK(v)dv = 0$, and $\int_{\mathbb{R}} |v|^3|\dot{K}(v)|dv < \infty$.

With these restrictions on the kernel function $K$, we obtain the following result for $\tilde{\mathbf{V}}_n^{\mathtt{MS}} = \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}})$, which improves slightly on Theorem A.3 because of the additional smoothness imposed. Let $F_{u|x_1,\mathbf{x}_2}^{(s)}(u|x_1,\mathbf{x}_2) = d^s F_{u|x_1,\mathbf{x}_2}(-u|u + x_1, x_2)/du^s$ and $f_{x_1|\mathbf{x}_2}^{(s)}(x_1|\mathbf{x}_2) = d^s f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)/dx_1^s$.

**Lemma MS** *Suppose Conditions MS and K(i)-(iii) hold.*

*If* $h_n \to 0$, $nh_n^3 \to \infty$, *and* $\mathbb{E}[\|\mathbf{x}_2\|^6] < \infty$, *then* $\tilde{\mathbf{V}}_n^{\mathtt{MS}} \to_{\mathbb{P}} \mathbf{V}_0^{\mathtt{MS}}$.

*Furthermore, if Condition K(iv) also holds,* $S \geq 4$, $S_F \geq 3$, *and* $S_f \geq 2$, *then*

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}} = \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) + O_{\mathbb{P}}(n^{-5/6}h_n^{-5/2}) + o_{\mathbb{P}}(h_n^2) + O_{\mathbb{P}}(n^{-1/3}),$$

*with the term* $O_{\mathbb{P}}(n^{-1/3})$ *independent of* $h_n$, *and*

$$\mathbb{E}\left[\|\tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) - \mathbf{V}_0^{\mathtt{MS}}\|^2\right] = h_n^4\Big(\sum_{k=1}^{d}\sum_{l=1}^{d}\mathsf{B}_{kl}^2\Big) + \frac{1}{nh_n^3}\Big(\sum_{k=1}^{d}\sum_{l=1}^{d}\mathsf{V}_{kl}\Big) + o(h_n^4 + n^{-1}h_n^{-3})$$

*where*

$$\begin{aligned}\mathsf{B}_{kl} = \mathbb{E}[(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)\{&F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}^{(2)}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\\ &+ F_{u|x_1,\mathbf{x}_2}^{(2)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\\ &+ \frac{1}{3}F_{u|x_1,\mathbf{x}_2}^{(3)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\}]\int v^3\dot{K}(v)dv\end{aligned}$$

*and*

$$\mathsf{V}_{kl} = \mathbb{E}[(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2 f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)]\int|\dot{K}(u)|^2du.$$

### A.4.3 Simulation Evidence

To investigate the finite sample properties of our proposed bootstrap-based inference procedures, we conducted a Monte Carlo experiment. Following Horowitz (2002), and to allow for a comparison with his bootstrap-based inference method for the smoothed maximum score estimator, we generate data from a model with $d = 1$, where

$$\mathbf{x} = (x_1, x_2)' \sim \mathcal{N}\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right),$$

and where $\varepsilon$ can take three distinct distributions. Specifically, DGP 1 sets $\varepsilon \sim \mathsf{Logistic}(0,1)/\sqrt{2\pi^2/3}$, DGP 2 sets $\varepsilon \sim \mathcal{T}_3/\sqrt{3}$, where $\mathcal{T}_k$ denotes a Student's t distribution with $k$ degrees of freedom, and DGP 3 sets $\varepsilon \sim (1 + 2(x_1 + x_2)^2 + (x_1 + x_2)^4)\mathsf{Logistic}(0,1)/\sqrt{\pi^2/48}$. The parameter is $\boldsymbol{\theta}_0 = 1$ in all cases.

The Monte Carlo experiment employs a sample size $n = 1,000$ with $B = 2,000$ bootstrap replications and $S = 2,000$ simulations. For each of the three DGPs, we implement the standard non-parametric bootstrap, $m$-out-of-$n$ bootstrap, and our proposed method using the two estimators $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ and $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ of $\mathbf{V}_0$. We report empirical coverage for nominal 95% confidence intervals and their average interval length. For the case of our proposed procedures, we investigate their performance using both (i) a grid of fixed tuning parameter values (bandwidth/derivative step) around the MSE-optimal choice and (ii) infeasible and feasible AMSE-optimal choices of the tuning parameter.

Table 1 presents the main results, which are consistent across all three simulation designs. First, as expected, the standard nonparametric bootstrap (labeled "Standard") does not perform well, leading to confidence intervals with an average 64% empirical coverage rate. Second, the $m$-out-of-$n$ bootstrap (labeled "m-out-of-n") performs somewhat better for small subsamples, but leads to very large average interval length of the resulting confidence intervals. Our proposed methods, on the other hand, exhibit excellent finite sample performance in this Monte Carlo experiment. Results employing the example-specific plug-in estimator $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ are presented under the label "Plug-in" while results employing the generic numerical derivative estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ are reported under the label "Num Deriv". Empirical coverage appears stable across different values of the tuning parameters for each method, with better performance in the case of $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$. We conjecture that $n = 1,000$ is too small for the numerical derivative estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ to lead to as good inference performance as $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ (e.g., note that the MSE-optimal choice $\epsilon_{\mathtt{MSE}}$ is greater than 1). Nevertheless, empirical coverage of confidence intervals constructed using our proposed bootstrap-based method is close to 95% in all cases except when $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ is used with either the infeasible asymptotic choice $\epsilon_{\mathtt{AMSE}}$ or its estimated counterpart $\hat{\epsilon}_{\mathtt{AMSE}}$, and with an average interval length of at most half that of any of the $m$-out-of-$n$ competing confidence intervals. In particular, confidence intervals based on $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ implemented with the feasible bandwidth $\hat{h}_{\mathtt{AMSE}}$ perform quite well across the three DGPs considered.

In sum, applying the bootstrap-based inference methods proposed in this note to the case of the Maximum Score estimator of Manski (1975) lead to confidence intervals with very good coverage and length properties in the simulation designs considered.

## A.5 Example: Conditional Maximum Score Estimation

Our results also apply directly to the conditional maximum score estimator introduced by Honoré and Kyriazidou (2000) in the context of a dynamic panel discrete choice model, which was also recently studied by Seo and Otsu (2018). The model is

$$Y_t = \mathbb{1}\{\mathbf{X}_t'\boldsymbol{\beta}_0 + \gamma_0 Y_{t-1} + \alpha + u_t \geq 0\}, \qquad t = 1, 2, 3,$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ and $\gamma_0 \in \mathbb{R}$ are unknown parameters of interest, and $\alpha$ is a time-invariant unobserved term. In addition to $\{Y_t, \mathbf{X}_t\}_{t=1}^3$, the initial condition $Y_0$ is observed. As in maximum score example, we need to impose a normalization on the parameters and we take $(\boldsymbol{\beta}_0', \gamma_0) = (1, \boldsymbol{\theta}_0')$ i.e the first element of $\boldsymbol{\beta}_0$ is unity.

To describe a version of the conditional maximum score estimator, partition $\mathbf{X}_t$ into $(X_{1t}, \mathbf{X}_{2t}')'$ where $X_{1t}$ is the first element of $\mathbf{X}_t$ and let $\mathbf{z} = (y, x_1, \mathbf{x}_2', \mathbf{w}')'$ where $y = Y_2 - Y_1$, $x_1 = X_{12} - X_{11}$, $\mathbf{x}_2 = ((\mathbf{X}_{22} - \mathbf{X}_{21})', Y_3 - Y_0)'$, and $\mathbf{w} = \mathbf{X}_2 - \mathbf{X}_3$. Then, the conditional maximum score estimator is an approximate maximizer $\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}}$ of $\hat{M}_n$ with

$$m_n(\mathbf{z}, \boldsymbol{\theta}) = m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta}) = y\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\} L_n(\mathbf{w})$$

where $b_n = o(1)$ is a sequence of bandwidths, and $L_n(\mathbf{x}) = b_n^{-d} L(\mathbf{x}/b_n)$ with $L(\cdot)$ a kernel function.

### A.5.1 Bootstrap-Based Inference

We employ the following additional notation:

$$\psi(x_1, \mathbf{x}_2, \mathbf{w}) = \{\mathbb{E}(A(Y_0, Y_3)|x_1, \mathbf{x}_2, \mathbf{w}) - \mathbb{E}(B(Y_0, Y_3)|x_1, \mathbf{x}_2, \mathbf{w})\} f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w}),$$

$$A(d_0, d_3) = \mathbb{1}(Y_0 = d_0, Y_1 = 0, Y_2 = 1, Y_3 = d_3),$$

$$B(d_0, d_3) = \mathbb{1}(Y_0 = d_0, Y_1 = 1, Y_2 = 0, Y_3 = d_3),$$

$$\varphi_1(\mathbf{w}; \boldsymbol{\theta}) = \int \int_{-\mathbf{x}_2' \boldsymbol{\theta}}^{\infty} \psi(x, \mathbf{x}_2, \mathbf{w}) dx dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}) f_{\mathbf{w}}(\mathbf{w}),$$

$$\varphi_2(\mathbf{w}) = \int \mathbf{x}_2 \psi(-\mathbf{x}_2' \boldsymbol{\theta}_0, \mathbf{x}_2, \mathbf{w}) dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}) f_{\mathbf{w}}(\mathbf{w}),$$

$$\mathbf{V}(\mathbf{w}; \boldsymbol{\theta}) = \int \mathbf{x}_2 \mathbf{x}_2' \left. \frac{d}{dx} \psi(x, \mathbf{x}_2, \mathbf{w}) \right|_{x = -\mathbf{x}_2' \boldsymbol{\theta}} dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}) f_{\mathbf{w}}(\mathbf{w}),$$

and $\mathbf{X} = \{\mathbf{X}_t\}_{t=1}^3$, $\mathcal{W} = \{\mathbf{w} : \|\mathbf{w}\| \leq \eta\}$, and $\mathcal{S} = \mathbb{R} \times \text{supp}(\mathbf{x}_2) \times \mathcal{W}$ for some $\eta > 0$. To verify Assumption CRA, we impose the following primitive set of conditions.

**Condition CMS** Let $S \geq 1$ and $S_f \geq 1$ be some integers.

(i) The parameter space $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\theta}_0$, which satisfies $(1, \boldsymbol{\theta}_0')' = (\boldsymbol{\beta}_0', \gamma_0)'$, lies in the interior of $\boldsymbol{\Theta}$.

(ii) The sample $\{\mathbf{z}_i\}_{i \geq 1}$ is i.i.d. across $i = 1, 2, \ldots, n$. The data generating process is

$$\mathbb{P}(Y_0 = 1|\mathbf{X}, \alpha) = p(\mathbf{X}, \alpha),$$

$$\mathbb{P}(Y_t = 1|\mathbf{X}, \alpha, Y_0, \ldots, Y_{t-1}) = F(\mathbf{X}_t' \boldsymbol{\beta}_0 + \gamma_0 Y_{t-1} + \alpha), \qquad t = 1, 2, 3,$$

where $F(\cdot)$ is strictly increasing on the entire real line.

(iii) $\mathbb{P}(y \neq 0|x_1, \mathbf{x}_2, \mathbf{w}) > 0$ on $\mathcal{S}$. The density $f_{\mathbf{w}}$ is continuous on $\mathcal{W}$ and $f_{\mathbf{w}}(\mathbf{0}) > 0$.

(iv) The support of $\mathbf{X}_1 - \mathbf{X}_2$ given $\mathbf{w} \equiv \mathbf{X}_2 - \mathbf{X}_3$ is not contained in any proper linear subspace of $\mathbb{R}^d$ for all $\mathbf{w} \in \mathcal{W}$ and $\mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}] \leq C$ for all $\mathbf{w} \in \mathcal{W}$.

(v) The conditional density $f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w})$ is positive and bounded on $\mathcal{S}$. In addition, there exists a function $B_f : \mathbb{R}^d \to \mathbb{R}_+$ such that, for some $\epsilon > 0$,

$$\sup_{\mathbf{w} \in \mathcal{W}} |f_{x_1}(\tilde{v}|\mathbf{x}_2, \mathbf{w}) - f_{x_1}(v|\mathbf{x}_2, \mathbf{w})| \leq B_f(v, \mathbf{x}_2)|\tilde{v} - v|^\epsilon,$$

with $\sup_{\mathbf{w} \in \mathcal{W}} \mathbb{E}[\|\mathbf{x}_2\|^3 B_f(-\mathbf{x}_2' \boldsymbol{\theta}_0, \mathbf{x}_2)|\mathbf{w}] < \infty$.

(vi) The function $\psi(x_1, \mathbf{x}_2, \mathbf{w})$ is bounded and $S$ times continuously differentiable in $x_1$ with bounded derivatives on $\mathcal{S}$.

(vii) The functions $\varphi_1(\mathbf{w};\boldsymbol{\theta})$ and $\mathbf{V}(\mathbf{w};\boldsymbol{\theta})$ satisfy, for some $\epsilon > 0$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\varphi_1(\mathbf{w};\boldsymbol{\theta}) - \varphi_1(\mathbf{0};\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta} \|\mathbf{V}(\mathbf{w};\boldsymbol{\theta}) - \mathbf{V}(\mathbf{0};\boldsymbol{\theta})\| \leq C\|\mathbf{w}\|^\epsilon, \qquad \mathbf{w} \in \mathcal{W}.$$

The function $\varphi_2$ is $S_f$ times continuously differentiable on $\mathcal{W}$. Also, $\mathbf{V}(\mathbf{0};\boldsymbol{\theta}_0)$ is positive definite.

(viii) The kernel function $L(\cdot)$ is bounded, compactly supported, $P$-th order with $1 \leq P \leq S_f$, and $\int L(\mathbf{v})d\mathbf{v} = 1$.

Under these conditions, we have the following result.

**Corollary CMS** *Suppose Condition CMS holds. If $nb_n^d \to \infty$ and $nb_n^{d+3P} \to 0$, then*

$$\sqrt[3]{nb_n^d}(\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}} - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d} \left\{ -\frac{1}{2}\mathbf{s}'\mathbf{V}_0^{\mathtt{CMS}}\mathbf{s} + \mathcal{G}_0(\mathbf{s}) \right\}$$

*where $\mathcal{G}_0$ is a mean-zero Gaussian process with covariance kernel $\mathcal{C}_0^{\mathtt{CMS}}(\cdot,\cdot)$, and*

$$\mathbf{V}_0^{\mathtt{CMS}} = \mathbb{E}[\mathbf{x}_2\mathbf{x}_2'\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2, \mathbf{0})|\mathbf{w} = \mathbf{0}]f_{\mathbf{w}}(\mathbf{0}), \qquad \psi^{(1)}(x, \mathbf{x}_2, \mathbf{w}) = \frac{d}{dx}\psi(x, \mathbf{x}_2, \mathbf{w}),$$

$$\mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s},\mathbf{t}) = \mathbb{E}\left[\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2, \mathbf{0})\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\} \,\middle|\, \mathbf{w} = \mathbf{0}\right] f_{\mathbf{w}}(\mathbf{0}).$$

*Furthermore, if a consistent estimator for the Hessian term is available (i.e., $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0^{\mathtt{CMS}}$ for some estimator $\tilde{\mathbf{V}}_n$), then the reshaped bootstrap consistently estimates the large sample distribution of $\sqrt[3]{nb_n^d}(\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}} - \boldsymbol{\theta}_0)$.*

### A.5.2 Drift Estimation

As we have done for maximum score example, we can also consider a case-specific estimator of $\mathbf{V}_0^{\mathtt{AMSE}}$. Our proposed estimator is

$$\tilde{\mathbf{V}}_n^{\mathtt{CMS}} = \tilde{\mathbf{V}}_n^{\mathtt{CMS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}}), \qquad \tilde{\mathbf{V}}_n^{\mathtt{CMS}}(\boldsymbol{\theta}) = -\frac{1}{nh_n^2}\sum_{i=1}^n y_i\dot{K}\left(\frac{x_{1i} + \mathbf{x}_{2i}'\boldsymbol{\theta}}{h_n}\right)\mathbf{x}_{2i}\mathbf{x}_{2i}'L_n(\mathbf{w}_i)$$

where $h_n = o(1)$ is a bandwidth, and $\dot{K}(u) = dK(u)/du$ with $K$ a kernel function.

We impose Condition K on the kernel function $K(\cdot)$, and obtain the following result for $\tilde{\mathbf{V}}_n^{\mathtt{CMS}}$. Let $\mathbf{V}_n^{\mathtt{CMS}} = \partial^2 M_n^{\mathtt{CMS}}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ where $M_n^{\mathtt{CMS}}(\boldsymbol{\theta}) = \mathbb{E}[m_n^{\mathtt{CMS}}(\mathbf{z},\boldsymbol{\theta})]$.

**Lemma CMS** *Suppose Condition CMS and Assumption K hold with $S \geq 2$.*

*If $h_n \to 0$ and $nb_n^d h_n^3 \to \infty$, and $\mathbb{E}[\|\mathbf{x}_2\|^6|\mathbf{w}] \leq C$ for $\mathbf{w} \in \mathcal{W}$, then $\tilde{\mathbf{V}}_n^{\mathtt{CMS}} \to_{\mathbb{P}} \mathbf{V}_0^{\mathtt{CMS}}$.*

*Furthermore, if $S \geq 3$, then*

$$\mathbb{E}\left[\|\tilde{\mathbf{V}}_n^{\mathtt{CMS}}(\boldsymbol{\theta}_0) - \mathbf{V}_n^{\mathtt{CMS}}\|^2\right] = O(h_n^4 + (nb_n^d h_n^3)^{-1}),$$

*where the rate is sharp in the sense that there exist constants* $\mathsf{B}_{kl}$ *and* $\mathsf{V}_{kl}$ *such that*

$$\frac{\mathbb{E}[(\tilde{V}^{\mathtt{CMS}}_{n,kl}(\boldsymbol{\theta}_0) - V^{\mathtt{CMS}}_{0,kl})^2]}{h_n^4\mathsf{B} + (nb_n^d h_n^3)^{-1}\mathsf{V}} \to 1,$$

*where* $\tilde{V}^{\mathtt{CMS}}_{n,kl}(\boldsymbol{\theta}_0)$ *and* $V^{\mathtt{CMS}}_{0,kl}$ *denote the* $(k,l)$*-th elements of* $\tilde{\mathbf{V}}^{\mathtt{CMS}}_n(\boldsymbol{\theta}_0)$ *and* $\mathbf{V}^{\mathtt{CMS}}_n$*, respectively.*

## A.6 Extension: Isotonic Density Estimation

Our main ideas and results can also be used to conduct valid bootstrap-based inference in the context of isotonic density estimation, another cube root consistent estimator with a Chernoff-type limiting distribution. See, e.g., van der Vaart and Wellner (1996, Example 3.2.14) for a modern textbook treatment of this example.

Suppose we observe $x_1, \ldots, x_n \overset{iid}{\sim} F$ where $F$ is continuous distribution function supported on $[0,\infty)$ with non-increasing density $f$. An extensively-studied method of estimating $f(x_0)$, $x_0 \in (0,\infty)$, is nonparametric maximum likelihood estimation (NPMLE):

$$\operatorname*{argmax}_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} \log f(x_i)$$

where $\mathcal{F}$ is the class of non-increasing densities supported on $[0,\infty)$. It is well known that the solution to this problem, denoted by $\hat{f}_n(x)$, is the left derivative at $x$ of the least concave majorant (LCM) of the empirical distribution function $\hat{F}_n(\cdot) = n^{-1}\sum_{i=1}^{n} \mathbb{1}(x_i \le \cdot)$.

The asymptotic distributional result for the NPMLE in this setting is as follows.

**Lemma ID** *If* $f$ *is differentiable at* $x_0$ *with* $f^{(1)}(x_0) < 0$*, then*

$$\sqrt[3]{n}(\hat{f}_n(x_0) - f(x_0)) \rightsquigarrow |4f^{(1)}(x_0)f(x_0)|^{1/3} \operatorname*{argmax}_{s\in\mathbb{R}} \left\{\mathcal{W}(s) - s^2\right\}$$

*where* $\mathcal{W}$ *is a standard two-sided Brownian motion with* $\mathcal{W}(0) = 0$*.*

The NPMLE estimator $\hat{f}_n(x_0)$ is not a member of the class of $M$-estimators considered in this paper, but nonetheless a distributional approximation of $\sqrt[3]{n}(\hat{f}_n(x_0) - f(x_0))$ can be established using the results for cube root consistent estimators discussed herein via the switching technique of Groeneboom (1985). Therefore, we first outline the main idea underlying this distributional approximation because it provides an insight to our proposed bootstrap-based inference approach.

The NPMLE estimator is the left-derivative of the least concave majorant (LCM) of $\hat{F}_n$ at $x_0$. Using this property, we have

$$\hat{S}_n(a) \le x \iff \hat{f}_n(x) \le a$$

where

$$\hat{S}_n(a) = \operatorname*{argmax}_{s\ge 0}\{\hat{F}_n(s) - as\}.$$

This is the switching technique alluded above. Then, we have

$$\mathbb{P}\left[n^{1/3}(\hat{f}_n(x_0) - f(x_0)) \le t\right] = \mathbb{P}\left[\hat{S}_n(f(x_0) + n^{-1/3}t) \le x_0\right].$$

Using the definition of $\hat{S}_n(a)$ and change-of-variable $x = x_0 + sn^{-1/3}$,

$$\begin{aligned}
\hat{S}_n(f(x_0) + n^{-1/3}t) &= \underset{x\in[0,\infty)}{\operatorname{argmax}} \left\{\hat{F}_n(x) - (f(x_0) + n^{-1/3}t)x\right\} \\
&= x_0 + n^{-1/3} \underset{s\in[-n^{1/3}x_0,\infty)}{\operatorname{argmax}} \left\{\hat{F}_n(x_0 + sn^{-1/3}) - (f(x_0) + n^{-1/3}t)(x_0 + sn^{-1/3})\right\} \\
&= x_0 + n^{-1/3} \underset{s\in[-n^{1/3}x_0,\infty)}{\operatorname{argmax}} \left\{\hat{F}_n(x_0 + sn^{-1/3}) - \hat{F}_n(x_0) - (f(x_0) + n^{-1/3}t)sn^{-1/3}\right\} \\
&= x_0 + n^{-1/3} \underset{s\in[-n^{1/3}x_0,\infty)}{\operatorname{argmax}} \left\{n^{2/3}(\hat{F}_n(x_0 + sn^{-1/3}) - \hat{F}_n(x_0) - f(x_0)sn^{-1/3}) - ts\right\}
\end{aligned}$$

where we use that a maximizer is invariant under shifting and scaling of the objective function.

Therefore, to study the limit law of $n^{1/3}(\hat{f}_n(x_0) - f(x_0))$, it suffices to look at the asymptotic distribution of the maximizer of $n^{2/3}(\hat{F}_n(x_0+sn^{-1/3}) - \hat{F}_n(x_0) - f(x_0)sn^{-1/3}) - ts$. To do that, we can first establish weak convergence of the objective function and then apply an argmax continuous mapping theorem.

To establish uniform weak convergence of the objective function, we let

$$n^{2/3}(\hat{F}_n(x_0 + sn^{-1/3}) - \hat{F}_n(x_0) - f(x_0)sn^{-1/3}) - ts = Q_n(s) + G_n(s)$$

where

$$\begin{aligned}
Q_n(s) &= n^{2/3}\left(F(x_0 + sn^{-1/3}) - F(x_0) - f(x_0)sn^{-1/3}\right) - ts, \\
G_n(s) &= n^{2/3}\left(\hat{F}_n(x_0 + sn^{-1/3}) - \hat{F}_n(x_0) - (F(x_0 + sn^{-1/3}) - F(x_0))\right).
\end{aligned}$$

For the first term, using differentiability of $f$ at $x_0$, a Taylor expansion gives

$$Q_n(s) = \frac{f^{(1)}(x_0)}{2}s^2 - ts + o(1),$$

which is non-random and the $o(1)$ term converges uniformly with respect to $s$ over a compact subset of $\mathbb{R}$. Note that in this $M$-estimation problem, the Hessian term $-V_0$ equals $f^{(1)}(x_0)$, the first derivative of $f$ at $x_0$.

For the second term, Lindeberg-Feller CLT implies finite-dimensional convergence and a standard maximal inequality shows stochastic equicontinuity, so we can establish that for any compact subset $\mathcal{S} \subset \mathbb{R}$,

$$G_n(s) \rightsquigarrow \mathcal{W}(f(x_0)s), \qquad s \in \mathcal{S}.$$

Then, an argmax continuous mapping theorem implies

$$\mathbb{P}\left(n^{1/3}(\hat{f}_n(x_0) - f(x_0)) \le t\right) \to \mathbb{P}\left(\underset{s}{\operatorname{argmax}}\left\{\mathcal{W}(f(x_0)s) + f^{(1)}(x_0)s^2/2 - ts\right\} \le 0\right).$$

Rewriting the object inside the argmax operator using properties of Brownian motion (c.f., van der Vaart and Wellner, 1996, Problem 3.2.5), we obtain the conclusion of Lemma ID.

### A.6.1 Bootstrap-based Inference

The above discussion indicates that, if $\tilde{\Delta}_n$ is the left-derivative at $t_0$ of the LCM of some generic function $U_n(x)$, we have

$$\mathbb{P}[n^{1/3}\tilde{\Delta}_n \le t] = \mathbb{P}\left(\underset{s}{\operatorname{argmax}}\left\{n^{2/3}(U_n(x_0 + sn^{-1/3}) - U_n(x_0)) - ts\right\} \le 0\right).$$

Therefore, for bootstrap validity, we need to study the bootstrap counterpart $n^{2/3}(U_n^*(x_0+sn^{-1/3}) - U_n^*(x_0))$ and, in particular, establish uniform weakly convergence in probability in order to establish validity of the bootstrap distributional approximation.

However, as discussed in our paper, in the cube root consistent case as in Kim and Pollard (1990) the bootstrap approximation of the objective function fails at the Hessian term. Therefore, we reshape the $M$-estimator as before prior to applying the bootstrap. In particular, let $\tilde{\Delta}_n^*$ be the left-derivative at $x = x_0$ of the LCM of $\tilde{F}_n^*(x)$, where

$$\tilde{F}_n^*(x) = \hat{F}_n^*(x) - \hat{F}_n(x) + \frac{1}{2}\tilde{V}_n(x - x_0)^2, \qquad \tilde{V}_n = \tilde{f}_n^{(1)}(x_0),$$

$\hat{F}_n^*(\cdot) = \sum_{i=1}^n \mathbb{1}(x_i^* \le \cdot)$, and $\tilde{f}_n^{(1)}(x_0)$ is a consistent estimator of $f^{(1)}(x_0)$.

The form of the above objective function exactly mimics the intuition described in the paper: we remove the problematic drift term under the bootstrap distribution and add back a plug-in estimate thereof.

**Theorem ID** *Suppose the conditions in Lemma ID hold. If $\tilde{f}_n^{(1)}(x_0) \to_{\mathbb{P}} f^{(1)}(x_0)$, then*

$$\sqrt[3]{n}\tilde{\Delta}_n^* \rightsquigarrow_{\mathbb{P}} |4f^{(1)}(x_0)f(x_0)|^{1/3} \underset{s\in\mathbb{R}}{\operatorname{argmax}}\left\{\mathcal{W}(s) - s^2\right\},$$

*where $\tilde{\Delta}_n^*$ be the left-derivative at $x = x_0$ of the LCM of $\tilde{F}_n^*(x) = \hat{F}_n^*(x) - \hat{F}_n(x) + \frac{1}{2}\tilde{V}_n(x - x_0)^2$.*

The above bootstrap-based construction employs the reshaped objective function $\tilde{F}_n^*(x)$. Another possibility is to use $\tilde{F}_n^*(x) + \hat{f}_n(x_0)x$ with $\hat{f}_n(x_0)$ being the NPMLE from the original sample. If we let $\tilde{f}_n^*(x_0)$ be the left-derivative at $x_0$ of the LCM of the alternative objective function, we have

$$\tilde{f}_n^*(x_0) - \hat{f}_n(x_0) = \tilde{\Delta}_n^*.$$

This follows because given $G(x) = g(x) + cx$ for some constant $c$, the LCM of $G$ equals the LCM of $g$ plus $cx$. In particular, suppose that this claim is false. Then, there exists a concave function $F$ such that $G(x) \leq F(x)$ for all $x$ and $F(\tilde{x}) < Lg(\tilde{x}) + c\tilde{x}$ for at least one $\tilde{x}$, where $Lg$ denotes the LCM of $g$. Thus, $F(\tilde{x}) - c\tilde{x} < Lg(\tilde{x})$ and this implies $F(\breve{x}) - c\breve{x} < g(\breve{x})$ for some $\breve{x}$, but this contradicts $G(x) \leq F(x)$ for all $x$.

### A.6.2 Drift Estimation

In this example the counterpart of $-\mathbf{V}_0$ is simply $f^{(1)}(x_0)$, the first derivative of the Lebesgue density $f(x)$ at $x_0$. Thus, the estimation of the drift term can be done using standard nonparametric estimation techniques, already available in most software platforms. For methodological and technical details see, for example, Wand and Jones (1995).

### A.6.3 Simulation Evidence

We investigate the finite sample properties of confidence intervals for $f(x_0)$ constructed using the bootstrap-based distributional approximation whose consistency was established in Theorem ID. We employ the DGPs and simulation setting previously considered in Sen, Banerjee, and Woodroofe (2010). This, as in the case of the Maximum Score estimator discussed in the paper, allows for a direct comparison with other bootstrap-based inference methods and their numerical performance already studied in previous work available in the literature.

We estimate $f(x_0)$ at the evaluation point $x_0 = 1$ using a random sample of observations, where three distinct distributions are considered: DGP 1 sets $x \sim \mathsf{Exponential}(1)$, DGP 2 sets $x \sim |\mathsf{Normal}(0, 1)|$, and DGP 3 sets $x \sim |\mathcal{T}_3|$. As in the case of the Maximum Score example, the Monte Carlo experiment employs a sample size $n = 1,000$ with $B = 2,000$ bootstrap replications and $S = 2,000$ simulations, and compares three types of bootstrap-based inference procedures: the standard non-parametric bootstrap, $m$-out-of-$n$ bootstrap, and our proposed method using two distinct estimators of $f'(x_0)$ (plug-in and numerical derivative).

Table 2 presents the numerical results for this example. We continue to report empirical coverage for nominal 95% confidence intervals and their average interval length. For the case of our proposed procedures, we again investigate their performance using both (i) a grid of fixed tuning parameter value (derivative step/bandwidth) and (ii) infeasible and feasible AMSE-optimal choice of tuning parameter. Also in this case, the numerical evidence is very encouraging. Our proposed bootstrap-based inference method leads to confidence intervals with excellent empirical coverage and average interval length, outperforming both the standard nonparametric bootstrap (which is inconsistent) and the $m$-out-of-$n$ bootstrap (which is consistent). In particular, in this example, the plug-in method employs an off-the-shelf kernel derivative estimator, which in this case leads to confidence intervals that are very robust (i.e., insensitive) to the choice of bandwidth. Furthermore, when the corresponding feasible off-the-shelf MSE-optimal bandwidth is used, the resulting confidence intervals continue to perform excellently. Finally, the generic numerical derivative estimator also leads to very good performance of bootstrap-based infeasible and feasible confidence intervals.

In sum, this example provides a second numerical illustration of the very good finite sample performance of inference based on our proposed bootstrap-based distributional approximation for cube root consistent estimators.

## A.7 Proofs

### A.7.1 Proof of Theorem A.1

Under the assumptions of the theorem,

$$r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{Q_n(\mathbf{s}) + \hat{G}_n(\mathbf{s})\} + o_{\mathbb{P}}(1),$$

where

$$Q_n(\mathbf{s}) = r_n^2[M_n(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) - M_n(\boldsymbol{\theta}_0)]\mathbb{1}(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1} \in \boldsymbol{\Theta})$$

and

$$\hat{G}_n(\mathbf{s}) = r_n^2[\hat{M}_n(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1}) + M_n(\boldsymbol{\theta}_0)]\mathbb{1}(\boldsymbol{\theta}_0 + \mathbf{s}r_n^{-1} \in \boldsymbol{\Theta}).$$

Moreover, $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ by Lemmas A.1 and A.3 and $Q_n + \hat{G}_n \rightsquigarrow \mathcal{Q}_0 + \mathcal{G}_0$ by Lemmas A.2, A.4, and A.5. The result now follows from the argmax continuous mapping theorem.

#### A.7.1.1 Consistency

**Lemma A.1** *Suppose Condition CRA(i) holds and suppose $r_n \to \infty$. If*

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

*then $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_{\mathbb{P}}(1)$.*

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

$$\mathbb{P}\left[\hat{M}_n(\boldsymbol{\theta}_0) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} \hat{M}_n(\boldsymbol{\theta}) > c_\delta\right] \to 1. \tag{A.2}$$

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

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{M}_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta})| = O_{\mathbb{P}}(\sqrt{n^{-1}\mathbb{E}[\bar{m}_n(\mathbf{z})^2]}) = O_{\mathbb{P}}(1/\sqrt{nq_n}) = O_{\mathbb{P}}(r_n^{-1/6}) = o_{\mathbb{P}}(1).$$

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

#### A.7.1.2 Local Behavior of $M_n$

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

$$\left| M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbf{V}_n(\boldsymbol{\theta}-\boldsymbol{\theta}_0) \right| \leq \dot{M}_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| + \frac{1}{2}\ddot{M}_n||\boldsymbol{\theta}-\boldsymbol{\theta}_0||^2, \qquad \text{(A.3)}$$

for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}_n$, where

$$\mathbf{V}_n = -\frac{1}{2}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_n(\boldsymbol{\theta}_0), \qquad \dot{M}_n = \left\|\frac{\partial}{\partial\boldsymbol{\theta}}M_n(\boldsymbol{\theta}_0)\right\|, \qquad \ddot{M}_n = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_n}\left\|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)]\right\|.$$

As an immediate consequence of (A.3), we have the following result about $Q_n$.

**Lemma A.2** *Suppose Condition CRA(ii) holds and suppose $r_n \to \infty$. Then, for any finite $K > 0$,*

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

**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.3) with $\boldsymbol{\Theta}_n = \boldsymbol{\Theta}_0^{Kr_n^{-1}}$, we have, uniformly in $\mathbf{s}$ with $||\mathbf{s}|| \leq K$,

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

where the last equality uses

$$\mathbf{V}_n - \mathbf{V}_0 = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta}_0) - M_0(\boldsymbol{\theta}_0)] \to 0, \qquad r_n\dot{M}_n = r_n\left\|\frac{\partial}{\partial\boldsymbol{\theta}}M_n(\boldsymbol{\theta}_0)\right\| \to 0,$$

and

$$\ddot{M}_n = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{Kr_n^{-1}}} \left\| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)] \right\|$$
$$\leq 2 \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{Kr_n^{-1}}} \left\| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})] \right\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{Kr_n^{-1}}} \left\| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_0(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta}_0)] \right\| \to 0,$$

which completes the proof. ∎

#### A.7.1.3 Rate of Convergence

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

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

*then* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_{\mathbb{P}}(r_n^{-1})$.

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

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

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 $C$ large.

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

$$\liminf_{n\to\infty}[\lambda_{\min}(\mathbf{V}_n) - \ddot{M}_n^\delta] > 0,$$

where $\lambda_{\min}(\cdot)$ denotes the minimal eigenvalue of the argument and

$$\mathbf{V}_n = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_n(\boldsymbol{\theta}_0), \qquad \ddot{M}_n^\delta = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta} \left\| \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}[M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)] \right\|.$$

Then, for all $n$ large enough and for any integer pair $(j, C)$ with $j \geq C$, we have:

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

where

$$c_n(\delta, C) = \frac{1}{8}[\lambda_{\min}(\mathbf{V}_n) - \ddot{M}_n^{\delta}] - 2^{-C} r_n \left\| \frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) \right\| - 2^{-2C}\delta,$$

and where the inequality uses the following implication of (A.3): If $\lambda_{\min}(\mathbf{V}_n) - \ddot{M}_n^{\delta} \geq 0$, then

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

for any subset $\boldsymbol{\Theta}_n'$ of $\boldsymbol{\Theta}_n$.

Choosing $C$ large enough that $\liminf_{n\to\infty} c_n(\delta, C) > 2c > 0$ for some $c > 0$, we may assume that $c_n(\delta, C) > c$ for every $n$, in which case

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

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

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

which can be made arbitrarily small by making $C$ large. ∎

#### A.7.1.4 Weak Convergence

To show that $\hat{G}_n \rightsquigarrow \mathcal{G}_0$, it suffices to show finite-dimensional convergence and stochastic equicontinuity.

**Lemma A.4** *Suppose Conditions CRA(iii) and CRA(iv) hold, $r_n \to \infty$, 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 and the Berry-Esseen inequality if it can be shown that, for any $\mathbf{s},\mathbf{t}\in\mathbb{R}^d$,

$$\mathbb{E}[\psi_n(\mathbf{z};\mathbf{s})\psi_n(\mathbf{z};\mathbf{t})] \to \mathcal{C}_0(\mathbf{s},\mathbf{t})$$

and

$$n^{-1/2}\mathbb{E}[|\psi_n(\mathbf{z};\mathbf{s})|^3] \to 0.$$

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

$$\psi_n(\mathbf{z};\mathbf{s}) = \sqrt{r_n q_n}[m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}r_n^{-1}) - m_n(\mathbf{z},\boldsymbol{\theta}_0)] - n^{-1/2}Q_n(\mathbf{s}),$$

we have:

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

and, using $\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^3] = o(q_n^{-2})$ (for $\delta_n = O(r_n^{-1})$),

$$\begin{aligned}
n^{-1/2}\mathbb{E}[|\psi(\mathbf{z};\mathbf{s})|^3] &\leq 8n^{-1/2}r_n^{3/2}q_n^{3/2}\mathbb{E}[|m_n(\mathbf{z},\boldsymbol{\theta}_0+\mathbf{s}r_n^{-1}) - m_n(\mathbf{z},\boldsymbol{\theta}_0)|^3] + 8n^{-2}|Q_n(\mathbf{s})|^3\\
&= o(n^{-1/2}r_n^{3/2}q_n^{3/2}q_n^{-2}) + o(n^{-1/2}) = o(1),
\end{aligned}$$

as was to be shown. ∎

**Lemma A.5** *Suppose Conditions CRA(iii) and CRA(v) hold and suppose* $r_n\to\infty$. *Then, for any finite* $K>0$ *and for any positive* $\Delta_n$ *with* $\Delta_n = o(1)$,

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

**Proof of Lemma A.5**. Let $K>0$ be given. Proceeding 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

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

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 $||\mathbf{s}||, ||\mathbf{t}|| \leq K$,

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

and therefore

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

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

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

which gives the result. ∎

### A.7.2 Proof of Theorem A.2

Under the assumptions of the theorem,

$$r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = \underset{\mathbf{s} \in \mathbb{R}^d}{\operatorname{argmax}} \{\tilde{Q}_n(\mathbf{s}) + \tilde{G}_n^*(\mathbf{s})\} + o_{\mathbb{P}}(1),$$

where

$$\tilde{Q}_n(\mathbf{s}) = r_n^2 [\tilde{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - \tilde{M}_n(\hat{\boldsymbol{\theta}}_n)] \mathbb{1}(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1} \in \boldsymbol{\Theta}) = -\frac{1}{2} \mathbf{s}' \tilde{\mathbf{V}}_n \mathbf{s} \mathbb{1}(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1} \in \boldsymbol{\Theta})$$

and

$$\tilde{G}_n^*(\mathbf{s}) = r_n^2 [\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - \tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \tilde{M}_n(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) + \tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n)] \mathbb{1}(\hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1} \in \boldsymbol{\Theta}).$$

Moreover, $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = O_{\mathbb{P}}(1)$ by Lemmas A.6 and A.8 and $\tilde{Q}_n + \tilde{G}_n^* \rightsquigarrow_{\mathbb{P}} \mathcal{Q}_0 + \mathcal{G}_0$ by Lemmas A.7, A.9, and A.10. The result now follows from the argmax continuous mapping theorem.

#### A.7.2.1 Consistency

**Lemma A.6** *Suppose Condition CRA(i) holds, $r_n \to \infty$, and suppose $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$, where $\mathbf{V}_0$ is positive definite. If*

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

*then $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(1)$.*

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

$$\mathbb{P}\left[\tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\hat{\boldsymbol{\Theta}}_n^\delta} \tilde{M}_n^*(\boldsymbol{\theta}) > c_\delta^*\right] \to 1, \tag{A.4}$$

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

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

It follows from Pollard (1989, Theorem 4.2) that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\hat{M}_n^*(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta})| = O_{\mathbb{P}}(\sqrt{n^{-1}\mathbb{E}[\bar{m}_n(\mathbf{z})^2]}) = O_{\mathbb{P}}(r_n^{-1/6}) = o_{\mathbb{P}}(1).$$

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

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

so (A.2) is satisfied with $c_\delta^* = \delta^2\lambda_{\min}(\mathbf{V}_0 + \mathbf{V}_0')/8$. ∎

#### A.7.2.2 Local Behavior of $\tilde{M}_n$

Because

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

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

**Lemma A.7** *Suppose $r_n \to \infty$, $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$, and suppose $\hat{\boldsymbol{\theta}}_n \to_{\mathbb{P}} \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$. Then, for any finite $K > 0$,*

$$\sup_{||\mathbf{s}||\leq K} |\tilde{Q}_n(\mathbf{s}) - \left(-\frac{1}{2}\mathbf{s}'\mathbf{V}_0\mathbf{s}\right)| \to_{\mathbb{P}} 0.$$

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

$$|\tilde{Q}_n(\mathbf{s}) - \left(-\frac{1}{2}\mathbf{s}'\mathbf{V}_0\mathbf{s}\right)| \leq \frac{1}{2}\left|\mathbf{s}'(\tilde{\mathbf{V}}_n - \mathbf{V}_0)\mathbf{s}\right| + \frac{1}{2}\left|\mathbf{s}'\mathbf{V}_0\mathbf{s}\right| \mathbb{1}(\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1} \notin \boldsymbol{\Theta}) \leq K^2 o_{\mathbb{P}}(1),$$

where the last equality uses $\tilde{\mathbf{V}}_n \to_{\mathbb{P}} \mathbf{V}_0$ and $\mathbb{P}(\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1} \notin \boldsymbol{\Theta}) \to 0$. ∎

#### A.7.2.3 Rate of Convergence

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

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

*then* $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n = O_{\mathbb{P}}(r_n^{-1})$.

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

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

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 $C$ 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 $C$ large.

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

$$\inf_{||\mathbf{V}-\mathbf{V}_0||\leq\delta} \lambda_{\min}(\mathbf{V} + \mathbf{V}') > \lambda_{\min}(\mathbf{V}_0 + \mathbf{V}_0')/2.$$

Then, if $||\tilde{\mathbf{V}}_n - \mathbf{V}_0|| \leq \delta$, we have, for any integer pair $(j, C)$ with $j \geq C$ :

$$\tilde{M}_n(\hat{\boldsymbol{\theta}}_n) - \sup_{2^{j-1}<r_n||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n||\leq 2^j} \tilde{M}_n(\boldsymbol{\theta}) - \delta r_n^{-2} \geq 2^{2j} c^*(\delta, C) r_n^{-2}$$

where

$$c^*(\delta, C) = \frac{1}{32}\lambda_{\min}(\mathbf{V}_0 + \mathbf{V}_0') - 2^{-2C}\delta.$$

Choosing $C$ large enough that $c^*(\delta, C) > c^* > 0$ for some $c^* > 0$ 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:

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

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 Pollard (1989, Theorem 4.2) can be used to show that the sum on the last line is bounded by a constant multiple of

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

which can be made arbitrarily small by making $C$ large. ∎

#### A.7.2.4 Weak Convergence

To show that $\tilde{G}_n^* \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0$, it suffices to show finite-dimensional conditional weak convergence in probability and stochastic equicontinuity.

**Lemma A.9** *Suppose Conditions CRA(iii) and CRA(iv) hold,* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_{\mathbb{P}}(r_n^{-1}) = o_{\mathbb{P}}(1)$, *and suppose* $\sup_{||\mathbf{s}||\leq K} |\hat{G}_n(\mathbf{s}) + Q_n(\mathbf{s})| = o_{\mathbb{P}}(\sqrt{n})$ *for every finite* $K > 0$. *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 and the Berry-Esseen inequality if it can be shown that, for any $\mathbf{s},\mathbf{t}\in\mathbb{R}^d$,

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

and

$$n^{-1/2}\mathbb{E}^*[|\hat{\psi}_n(\mathbf{z}^*;\mathbf{s})|^3]=\frac{1}{n^{3/2}}\sum_{i=1}^{n}|\hat{\psi}_n(\mathbf{z}_i;\mathbf{s})|^3\to_{\mathbb{P}}0.$$

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),

$$\begin{aligned}&r_nq_n\mathbb{E}^*[\{m_n(\mathbf{z}^*,\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-m_n(\mathbf{z}^*,\hat{\boldsymbol{\theta}}_n)\}\{m_n(\mathbf{z}^*,\hat{\boldsymbol{\theta}}_n+\mathbf{t}r_n^{-1})-m_n(\mathbf{z}^*,\hat{\boldsymbol{\theta}}_n)\}]-\hat{\mathcal{C}}_n(\mathbf{s},\mathbf{t})\\ =\ &r_nq_n\frac{1}{n}\sum_{i=1}^{n}\{m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n)\}\{m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n+\mathbf{t}r_n^{-1})-m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n)\}-\hat{\mathcal{C}}_n(\mathbf{s},\mathbf{t})\\ =\ &o_{\mathbb{P}}(r_nq_nn^{-1/2}\sqrt{q_n^{-3}r_n})=o_{\mathbb{P}}(1),\end{aligned}$$

where

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

Using these facts and the representation

$$\hat{\psi}_n(\mathbf{z};\mathbf{s})=\sqrt{r_nq_n}[m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-m_n(\mathbf{z},\hat{\boldsymbol{\theta}}_n)]-n^{-1/2}\hat{Q}_n(\mathbf{s}),$$

we have:

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

and, using $\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^3] = o(q_n^{-2})$ (for $\delta_n = O(r_n^{-1})$),

$$\begin{aligned}
n^{-1/2}\mathbb{E}^*[|\hat{\psi}_n(\mathbf{z}^*;\mathbf{s})|^3] &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}|\hat{\psi}_n(\mathbf{z}_i;\mathbf{s})|^3 \\
&\leq 8n^{-1/2}r_n^{3/2}q_n^{3/2}\frac{1}{n}\sum_{i=1}^{n}|m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n+\mathbf{s}r_n^{-1})-m_n(\mathbf{z}_i,\hat{\boldsymbol{\theta}}_n)|^3+8n^{-2}|\hat{Q}_n(\mathbf{s})|^3 \\
&= o_{\mathbb{P}}(n^{-1/2}r_n^{3/2}q_n^{3/2}q_n^{-2})+o_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(1),
\end{aligned}$$

as was to be shown. ∎

**Lemma A.10** *Suppose Conditions CRA(iii) and CRA(v) hold and suppose* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_{\mathbb{P}}(r_n^{-1}) = o_{\mathbb{P}}(1)$. *Then, for any finite* $K>0$ *and for any positive* $\Delta_n$ *with* $\Delta_n = o(1)$,

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

**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 $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_{\mathbb{P}}(r_n^{-1})$, it suffices to show that, for every finite $k>0$,

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

where

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

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

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

and therefore

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

$$+ C r_n \sup_{\substack{||\mathbf{s}-\mathbf{t}||\leq\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\leq K+k}} \mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]$$

$$+ C r_n \mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}||\leq\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\leq K+k}} |\frac{1}{n}\sum_{i=1}^{n}\{d_n(\mathbf{z}_{i,n};\mathbf{s},\mathbf{t}) - \mathbb{E}[d_n(\mathbf{z};\mathbf{s},\mathbf{t})]\}|\right]$$

$$+ C r_n \mathbb{E}\left[\sup_{\substack{||\mathbf{s}-\mathbf{t}||\leq\Delta_n\\ ||\mathbf{s}||,||\mathbf{t}||\leq K+k}} |\frac{1}{n}\sum_{i=1}^{n}\{d_n(\mathbf{z}^*_{i,n};\mathbf{s},\mathbf{t}) - \mathbb{E}^*[d_n(\mathbf{z}^*;\mathbf{s},\mathbf{t})]\}|\right].$$

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

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

which gives the result. ∎

### A.7.3 Proof of Theorem A.3

#### A.7.3.1 Consistency

Without loss of generality, assume $\mathbb{P}[r_n|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0| \leq C] = 1$ for some fixed constant $C$. Write $\bar{V}_{n,kl}(\boldsymbol{\theta}) = \mathbb{E}[\tilde{V}_{n,kl}(\boldsymbol{\theta})]$. Consider the following decomposition

$$\tilde{V}^{\text{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) = \bar{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0) + \left[\tilde{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0) - \bar{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0)\right] + R_{1,n} + R_{2,n},$$

where

$$R_{1,n} = \bar{V}^{\text{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) - \bar{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0)$$
$$R_{2,n} = \tilde{V}^{\text{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) - \tilde{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0) - \bar{V}^{\text{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n) + \bar{V}^{\text{ND}}_{n,kl}(\boldsymbol{\theta}_0).$$

By definition,

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

Using (A.3),

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

where we use $\dot{M}_n = o(r_n^{-1})$ and $\ddot{M}_n = o(1)$. Then,

$$\begin{aligned} \bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) &= \frac{1}{4}\left[(\mathbf{e}_k + \mathbf{e}_l)'\mathbf{V}_n(\mathbf{e}_k + \mathbf{e}_l) - (\mathbf{e}_k - \mathbf{e}_l)'\mathbf{V}_n(\mathbf{e}_k - \mathbf{e}_l)\right] + o(r_n^{-1}\epsilon_n^{-1} + 1) \\ &= \mathbf{e}_k' \mathbf{V}_n \mathbf{e}_l + o(r_n^{-1}\epsilon_n^{-1} + 1). \end{aligned}$$

By condition CRA(ii), we have $\mathbf{V}_n \to \mathbf{V}_0$. Therefore, $\bar{\mathbf{V}}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0)$ converges to $\mathbf{e}_k' \mathbf{V}_0 \mathbf{e}_l = V_{0,kl}$.

It remains to show $\tilde{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) - \bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) = o_{\mathbb{P}}(1)$, $R_{1,n} = o_{\mathbb{P}}(1)$, and $R_{2,n} = o_{\mathbb{P}}(1)$. First, note that $\tilde{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) - \bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0)$ is mean zero and its variance is bounded by a constant multiple of $\epsilon_n^{-4} n^{-1}\mathbb{E}[\bar{d}_n^{2\epsilon_n}(\mathbf{z})^2] = O(n^{-1}q_n^{-1}\epsilon_n^{-3})$. Then, $\tilde{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) - \bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0) = O_{\mathbb{P}}((\epsilon_n r_n)^{-3/2}) = o_{\mathbb{P}}(1)$.

Next, using (A.3),

$$\begin{aligned} & M_n(\hat{\boldsymbol{\theta}}_n + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \\ & = -\frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])'\mathbf{V}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) + \frac{\epsilon_n^2}{2}(\mathbf{e}_k + \mathbf{e}_l)'\mathbf{V}_n(\mathbf{e}_k + \mathbf{e}_l) + o(r_n^{-1}\epsilon_n + \epsilon_n^2) \\ & \leq C(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 + \epsilon_n\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|) + o(r_n^{-1}\epsilon_n + \epsilon_n^2), \end{aligned}$$

and therefore $R_{1,n} = O_{\mathbb{P}}(r_n^{-2}\epsilon_n^{-2} + \epsilon_n^{-1}r_n^{-1}) + o(r_n^{-1}\epsilon_n^{-1} + 1) = o_{\mathbb{P}}(1)$.

Finally, by adding and subtracting $\hat{M}_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0)$ to the left-hand side of $R_{2,n}$, it suffices to analyze

$$\epsilon_n^{-2} \sup_{|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq C r_n^{-1} + 2\epsilon_n} \left| \hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - \{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)\} \right|.$$

Letting $t_n = Cr_n^{-1} + 2\epsilon_n$, and applying the maximal inequality in Pollard (1989, Theorem 4.2),

$$\begin{aligned} & \epsilon_n^{-2}\mathbb{E} \sup_{|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \leq t_n} \left| \hat{M}_n(\boldsymbol{\theta}) - \hat{M}_n(\boldsymbol{\theta}_0) - \{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)\} \right| \\ & \leq C\epsilon_n^{-2} n^{-1/2} (\mathbb{E}|\bar{d}_n^{t_n}(\mathbf{z})|^2)^{1/2} \\ & \leq C\epsilon_n^{-2} n^{-1/2} q_n^{-1/2} (r_n^{-1} + \epsilon_n)^{1/2} = O((\epsilon_n r_n)^{-3/2}). \end{aligned}$$

This implies that $R_{2,n} = o_{\mathbb{P}}(1)$.

#### A.7.3.2 Approximate Mean Squared Error

Define $\check{\mathbf{V}}_n^{\texttt{ND}} = \tilde{\mathbf{V}}_n^{\texttt{ND}}(\boldsymbol{\theta}_0)$. Since $\tilde{V}_{n,kl}^{\texttt{ND}} = \check{V}_{n,kl}^{\texttt{ND}} + R_{1,n} + R_{2,n}$, our first goal is to show that $R_{1,n} = o_{\mathbb{P}}(\epsilon_n^2) + O_{\mathbb{P}}(r_n^{-1})$ and $R_{2,n} = o_{\mathbb{P}}(r_n^{-3/2}\epsilon_n^{-3/2})$. For $\boldsymbol{\theta}, \boldsymbol{\vartheta} \in \boldsymbol{\Theta}_0^{\delta}$, using a Taylor approximation,

$$\begin{aligned}
&M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\vartheta})\\
&= (\boldsymbol{\theta} - \boldsymbol{\vartheta})' \frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\vartheta}) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\vartheta})' \mathbf{V}_n (\boldsymbol{\theta} - \boldsymbol{\vartheta})\\
&\quad + \frac{1}{6} \sum_{j_1, j_2, j_3} \frac{\partial^3 M_n(\boldsymbol{\vartheta})}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} (\theta_{j_1} - \vartheta_{j_1})(\theta_{j_2} - \vartheta_{j_2})(\theta_{j_3} - \vartheta_{j_3})\\
&\quad + \frac{1}{24} \sum_{j_1, j_2, j_3, j_4} \frac{\partial^4 M_n(\boldsymbol{\vartheta} + \boldsymbol{\xi})}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3} \partial \theta_{j_4}} (\theta_{j_1} - \vartheta_{j_1})(\theta_{j_2} - \vartheta_{j_2})(\theta_{j_3} - \vartheta_{j_3})(\theta_{j_4} - \vartheta_{j_4}) \qquad \text{(A.5)}
\end{aligned}$$

where $\boldsymbol{\xi}$ is a mean-value between $\mathbf{0}$ and $\boldsymbol{\theta} - \boldsymbol{\vartheta}$. Then,

$$\begin{aligned}
&M_n(\hat{\boldsymbol{\theta}}_n + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) + M_n(\boldsymbol{\theta}_0)\\
&= (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \frac{\partial}{\partial \boldsymbol{\theta}} M_n(\boldsymbol{\theta}_0) - \frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)' \mathbf{V}_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) - \epsilon_n (\mathbf{e}_k + \mathbf{e}_l)' \mathbf{V}_n (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\
&\quad + \frac{1}{6} \sum_{j_1, j_2, j_3} \frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} \mathbf{e}_{j_1}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\
&\quad + \frac{\epsilon_n}{2} \sum_{j_1, j_2, j_3} \frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\
&\quad + \frac{\epsilon_n^2}{2} \sum_{j_1, j_2, j_3} \frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\
&\quad + \frac{\epsilon_n^3}{6} \sum_{j_1, j_2, j_3} \frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_3}'(\mathbf{e}_k + \mathbf{e}_l)\\
&\quad + \frac{1}{24} \sum_{j_1, j_2, j_3, j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0 + \boldsymbol{\xi}_1)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3} \partial \theta_{j_4}} \mathbf{e}_{j_1}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\\
&\qquad\qquad \times \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\\
&\quad - \frac{\epsilon_n^4}{24} \sum_{j_1, j_2, j_3, j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0 + \boldsymbol{\xi}_2)}{\partial \theta_{j_1} \partial \theta_{j_2} \partial \theta_{j_3} \partial \theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_3}'(\mathbf{e}_k + \mathbf{e}_l) \mathbf{e}_{j_4}'(\mathbf{e}_k + \mathbf{e}_l)
\end{aligned}$$

where $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are also mid-points as $\boldsymbol{\xi}$. Using uniform convergence of fourth derivatives, the last two lines equal

$$\begin{aligned}
&= \frac{1}{24} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \\
&\qquad \times \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \\
&\quad - \frac{\epsilon_n^4}{24} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_3}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_4}'(\mathbf{e}_k + \mathbf{e}_l) + o_{\mathbb{P}}(\epsilon_n^4)
\end{aligned}$$

where we use $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| = o_{\mathbb{P}}(\epsilon_n)$. By expanding the terms,

$$\begin{aligned}
&\frac{1}{24} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \\
&\qquad \times \mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l])\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) \\
&= \frac{1}{24} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad + \frac{\epsilon_n}{6} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_2}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad + \frac{\epsilon_n^2}{2} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_3}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad + \frac{\epsilon_n^3}{4} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_3}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_4}'(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\quad + \frac{\epsilon_n^4}{24} \sum_{j_1,j_2,j_3,j_4} \frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}} \mathbf{e}_{j_1}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_2}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_3}'(\mathbf{e}_k + \mathbf{e}_l)\mathbf{e}_{j_4}'(\mathbf{e}_k + \mathbf{e}_l).
\end{aligned}$$

Then, because

$$\begin{aligned}
&-4\epsilon_n^2 \left[\bar{V}_{n,kl}^{\mathtt{ND}}(\hat{\boldsymbol{\theta}}_n) - \bar{V}_{n,kl}^{\mathtt{ND}}(\boldsymbol{\theta}_0)\right] \\
&= M_n(\hat{\boldsymbol{\theta}}_n + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) + M_n(\boldsymbol{\theta}_0) \\
&\quad - M_n(\hat{\boldsymbol{\theta}}_n + \epsilon_n[\mathbf{e}_k - \mathbf{e}_l]) + M_n(\boldsymbol{\theta}_0) + M_n(\boldsymbol{\theta}_0 + \epsilon_n[\mathbf{e}_k - \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0) \\
&\quad - M_n(\hat{\boldsymbol{\theta}}_n - \epsilon_n[\mathbf{e}_k - \mathbf{e}_l]) + M_n(\boldsymbol{\theta}_0) + M_n(\boldsymbol{\theta}_0 - \epsilon_n[\mathbf{e}_k - \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0) \\
&\quad + M_n(\hat{\boldsymbol{\theta}}_n - \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) - M_n(\boldsymbol{\theta}_0) - M_n(\boldsymbol{\theta}_0 - \epsilon_n[\mathbf{e}_k + \mathbf{e}_l]) + M_n(\boldsymbol{\theta}_0),
\end{aligned}$$

we have

$$
\begin{aligned}
&-4\epsilon_n^2\left[\bar{V}^{\mathtt{ND}}_{n,kl}(\hat{\boldsymbol{\theta}}_n)-\bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0)\right]\\
&=\epsilon_n^2\sum_{j_1,j_2,j_3}\frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\mathbf{e}'_{j_1}(\mathbf{e}_k+\mathbf{e}_l)\mathbf{e}'_{j_2}(\mathbf{e}_k+\mathbf{e}_l)\mathbf{e}'_{j_3}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\
&\quad-\epsilon_n^2\sum_{j_1,j_2,j_3}\frac{\partial^3 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}}\mathbf{e}'_{j_1}(\mathbf{e}_k-\mathbf{e}_l)\mathbf{e}'_{j_2}(\mathbf{e}_k-\mathbf{e}_l)\mathbf{e}'_{j_3}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\
&\quad+\epsilon_n^2\sum_{j_1,j_2,j_3,j_4}\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}}\mathbf{e}'_{j_1}(\mathbf{e}_k+\mathbf{e}_l)\mathbf{e}'_{j_2}(\mathbf{e}_k+\mathbf{e}_l)\mathbf{e}'_{j_3}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\mathbf{e}'_{j_4}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\\
&\quad-\epsilon_n^2\sum_{j_1,j_2,j_3,j_4}\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_{j_1}\partial\theta_{j_2}\partial\theta_{j_3}\partial\theta_{j_4}}\mathbf{e}'_{j_1}(\mathbf{e}_k-\mathbf{e}_l)\mathbf{e}'_{j_2}(\mathbf{e}_k-\mathbf{e}_l)\mathbf{e}'_{j_3}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\mathbf{e}'_{j_4}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o_{\mathbb{P}}(\epsilon_n^4)\\
&=O_{\mathbb{P}}(\epsilon_n^2 r_n^{-1})+o_{\mathbb{P}}(\epsilon_n^4)
\end{aligned}
$$

where we use $r_n^{-1}=o(\epsilon_n)$. Note that the term $O_{\mathbb{P}}(\epsilon_n^2 r_n^{-1})$ is independent of $\epsilon_n$ after divided by $\epsilon_n^2$. Therefore, $R_{1,n}=o_{\mathbb{P}}(\epsilon_n^2)+O_{\mathbb{P}}(r_n^{-1})$.

For term $R_{2,n}$, consider

$$
\epsilon_n^{-2}\left[\hat{M}_n(\hat{\boldsymbol{\theta}}_n+\epsilon_n\mathbf{e})-\hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{e})-M_n(\hat{\boldsymbol{\theta}}_n+\epsilon_n\mathbf{e})+M_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{e})\right]
$$

for some $d$-dimensional vector $\mathbf{e}$. It suffices to show

$$
\epsilon_n^{-2}\sup_{\substack{\|\mathbf{s}-\mathbf{t}\|\le\Delta_n\\ \|\mathbf{s}\|,\|\mathbf{t}\|\le C}}\left|\hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{s})-\hat{M}_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{t})-M_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{s})+M_n(\boldsymbol{\theta}_0+\epsilon_n\mathbf{t})\right|=o_{\mathbb{P}}(r_n^{-3/2}\epsilon_n^{-3/2})
$$

for $\Delta_n=o(1)$. Note that this is almost identical to the statement in Lemma A.8, except that $r_n$ in Lemma A.8 is replaced by $\epsilon_n^{-1}$. Doing the same calculation,

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

Now, we calculate the constants for leading bias and variance. Using (A.5),

$$
\begin{aligned}
\mathbb{E}[\check{V}^{\mathtt{ND}}_{n,kl}]&=\bar{V}^{\mathtt{ND}}_{n,kl}(\boldsymbol{\theta}_0)\\
&=\frac{1}{2}\mathbf{e}'_k\mathbf{V}_n\mathbf{e}_l+\frac{\epsilon_n^2}{6}\left(\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_k^3\partial\theta_l}+\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_k\partial\theta_l^3}\right).
\end{aligned}
$$

Thus, the constant for the leading bias is

$$
\mathsf{B}_{kl}=\frac{1}{6}\lim_{n\to\infty}\left(\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_k^3\partial\theta_l}+\frac{\partial^4 M_n(\boldsymbol{\theta}_0)}{\partial\theta_k\partial\theta_l^3}\right)=\frac{1}{6}\left(\frac{\partial^4 M_0(\boldsymbol{\theta}_0)}{\partial\theta_k^3\partial\theta_l}+\frac{\partial^4 M_0(\boldsymbol{\theta}_0)}{\partial\theta_k\partial\theta_l^3}\right).
$$

Next, consider the variance of $\check{V}^{\texttt{ND}}_{n,kl}$. To save space, write

$$m_{ni}(\epsilon_1, \epsilon_2) = m_n(\mathbf{z}_i, \boldsymbol{\theta}_0 + \epsilon_1 \mathbf{e}_k + \epsilon_2 \mathbf{e}_l)$$

Recall that

$$\check{V}^{\texttt{ND}}_{n,kl} = -\frac{1}{4\epsilon_n^2 n} \sum_{i=1}^{n} [m_{ni}(\epsilon_n, \epsilon_n) - m_{ni}(\epsilon_n, -\epsilon_n) - m_{ni}(-\epsilon_n, \epsilon_n) + m_{ni}(-\epsilon_n, -\epsilon_n)]$$

and

$$\begin{aligned}
\mathbb{V}[\tilde{V}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0)] &= \frac{1}{16\epsilon_n^4 n} \mathbb{V}\left(m_{ni}(\epsilon_n, \epsilon_n) - m_{ni}(\epsilon_n, -\epsilon_n) - m_{ni}(-\epsilon_n, \epsilon_n) + m_{ni}(-\epsilon_n, -\epsilon_n)\right) \\
&= \frac{1}{16\epsilon_n^4 n} \mathbb{E}\left[\{m_{ni}(\epsilon_n, \epsilon_n) - m_{ni}(\epsilon_n, -\epsilon_n) - m_{ni}(-\epsilon_n, \epsilon_n) + m_{ni}(-\epsilon_n, -\epsilon_n)\}^2\right] \\
&\quad - \frac{1}{n}[\bar{\mathbf{V}}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0)]^2
\end{aligned}$$

From the above calculation, the second term in the last equation is $O(n^{-1})$. By condition CRA(iv)

$$\epsilon_n^{-1} q_n \mathbb{E}[\{m_n(\mathbf{z}, \boldsymbol{\theta}_0 + \mathbf{s}\epsilon_n) - m_n(\mathbf{z}, \boldsymbol{\theta}_0)\}\{m_n(\mathbf{z}, \boldsymbol{\theta}_0 + \mathbf{t}\epsilon_n) - m_n(\mathbf{z}, \boldsymbol{\theta}_0)\}] \to \mathcal{C}_0(\mathbf{s}, \mathbf{t}).$$

We have

$$\begin{aligned}
&\lim_{n\to\infty} \epsilon_n^{-1} q_n \mathbb{E}\left[\{m_{ni}(\epsilon_n, \epsilon_n) - m_{ni}(\epsilon_n, -\epsilon_n) - m_{ni}(-\epsilon_n, \epsilon_n) + m_{ni}(-\epsilon_n, -\epsilon_n)\}^2\right] \\
&= 2\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, \mathbf{e}_k + \mathbf{e}_l) + 2\mathcal{C}_0(\mathbf{e}_k - \mathbf{e}_l, \mathbf{e}_k - \mathbf{e}_l) - 4\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, \mathbf{e}_k - \mathbf{e}_l) - 4\mathcal{C}_0(\mathbf{e}_k + \mathbf{e}_l, -\mathbf{e}_k + \mathbf{e}_l)
\end{aligned}$$

where we use $\mathcal{C}_0(\mathbf{s}, -\mathbf{s}) = 0$ and $\mathcal{C}_0(\mathbf{s}, \mathbf{t}) = \mathcal{C}_0(-\mathbf{s}, -\mathbf{t})$. Then,

$$\mathbb{V}[\tilde{V}^{\texttt{ND}}_{n,kl}(\boldsymbol{\theta}_0)] = \frac{1}{r_n^3 \epsilon_n^3}[\mathsf{V}_{kl} + o(1)] + O(n^{-1})$$

where $\mathsf{V}_{kl} = \frac{1}{8}[\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l, \mathbf{e}_k+\mathbf{e}_l) + \mathcal{C}_0(\mathbf{e}_k-\mathbf{e}_l, \mathbf{e}_k-\mathbf{e}_l) - 2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l, \mathbf{e}_k-\mathbf{e}_l) - 2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l, -\mathbf{e}_k+\mathbf{e}_l)]$.

### A.7.4 Proof of Corollary MS

It suffices to verify that Condition MS implies Condition $\text{CRA}_0$.

*Condition $CRA_0$(i).* The manageability assumption can be verified using the same argument as in Kim and Pollard (1990). Note that the function $|m^{\texttt{MS}}(\mathbf{z}, \boldsymbol{\theta})|$ is bounded by unity in this example, and thus finite second moment condition holds. Arguing as in Manski (1985), we can 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) and MS(iv) imply this condition.

*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 $d_0^\delta(\mathbf{z})^2 = d_0^\delta(\mathbf{z})$, $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^3] = O(\delta_n)$ and $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^4] = O(\delta_n)$, which implies the first condition. As noted in Abrevaya and Huang (2005),

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

From this representation, it follows that $\mathcal{C}_0^{\mathtt{MS}}(\mathbf{s},\mathbf{s}) + \mathcal{C}_0^{\mathtt{MS}}(\mathbf{t},\mathbf{t}) - 2\mathcal{C}_0^{\mathtt{MS}}(\mathbf{s},\mathbf{t}) > 0$ for $\mathbf{s} \neq \mathbf{t}$. Using $2xy = x^2 + y^2 - (x-y)^2$, it suffices to show for $\delta_n = O(r_n^{-1})$,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}} \left| r_n \mathbb{E}|m(\mathbf{z}_1, \boldsymbol{\theta} + \mathbf{s}/r_n) - m(\mathbf{z}_1, \boldsymbol{\theta} + \mathbf{t}/r_n)|^2 - \mathcal{L}_0(\mathbf{s}-\mathbf{t}) \right| = o(1)$$

for $\mathcal{L}_0$ defined in Corollary MS. We have

$$\begin{aligned}
&\mathbb{E}|m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}_1) - m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta}_2)|^2 \\
&= \mathbb{E}\mathbb{1}(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_1 \geq 0 > x_1 + \mathbf{x}_2'\boldsymbol{\theta}_2) + \mathbb{E}\mathbb{1}\left(x_1 + \mathbf{x}_2'\boldsymbol{\theta}_2 \geq 0 > x_1 + \mathbf{x}_2'\boldsymbol{\theta}_1\right) \\
&= \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_2|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_1|\mathbf{x}_2)\}\mathbb{1}\left(\mathbf{x}_2'\boldsymbol{\theta}_2 < \mathbf{x}_2'\boldsymbol{\theta}_1\right)\right] \\
&\quad + \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\boldsymbol{\theta}_1'\mathbf{x}_2|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_2|\mathbf{x}_2)\}\mathbb{1}\left(\mathbf{x}_2'\boldsymbol{\theta}_1 < \mathbf{x}_2'\boldsymbol{\theta}_2\right)\right] \\
&= \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\mathbb{1}\left(\mathbf{x}_2'\boldsymbol{\theta}_2 < \mathbf{x}_2'\boldsymbol{\theta}_1\right)] \\
&\quad + \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)\mathbb{1}\left(\mathbf{x}_2'\boldsymbol{\theta}_1 < \mathbf{x}_2'\boldsymbol{\theta}_2\right)] + R
\end{aligned}$$

where $\boldsymbol{\theta}_1 = \boldsymbol{\theta} + \mathbf{s}/r_n$, $\boldsymbol{\theta}_2 = \boldsymbol{\theta} + \mathbf{t}/r_n$, and

$$\begin{aligned}
R = &\ \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_2|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) - f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\boldsymbol{\theta}_2 + \boldsymbol{\theta}_0)\}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta}_2 < \mathbf{x}_2'\boldsymbol{\theta}_1)\right] \\
&- \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_1|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) - f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\boldsymbol{\theta}_1 + \boldsymbol{\theta}_0)\}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta}_2 < \mathbf{x}_2'\boldsymbol{\theta}_1)\right] \\
&+ \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_1|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) - f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\boldsymbol{\theta}_1 + \boldsymbol{\theta}_0)\}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta}_1 < \mathbf{x}_2'\boldsymbol{\theta}_2)\right] \\
&- \mathbb{E}\left[\{F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_2|\mathbf{x}_2) - F_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) - f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\boldsymbol{\theta}_2 + \boldsymbol{\theta}_0)\}\mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta}_1 < \mathbf{x}_2'\boldsymbol{\theta}_2)\right].
\end{aligned}$$

By continuous differentiability of $f_{x_1|\mathbf{x}_2}$ with bounded derivative,

$$|R| \leq 4 \sup_{x_1,\mathbf{x}_2} \left| f_{x_1|\mathbf{x}_2}^{(1)}(x_1|\mathbf{x}_2) \right| \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}} \|\boldsymbol{\theta} - \boldsymbol{\theta}_0 + (\mathbf{s}+\mathbf{t})r_n^{-1}\|^2 \mathbb{E}\|\mathbf{x}_2\|^2 = o(r_n^{-1}).$$

Then,

$$\begin{aligned}
&r_n\mathbb{E}|m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta} + \mathbf{s}/r_n) - m^{\mathtt{MS}}(\mathbf{z},\boldsymbol{\theta} + \mathbf{t}/r_n)|^2 \\
&= \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\mathbf{t}+\mathbf{s})\mathbb{1}(\mathbf{x}_2'\mathbf{t} < \mathbf{x}_2'\mathbf{s})] + \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2'(-\mathbf{s}+\mathbf{t})\mathbb{1}(\mathbf{x}_2'\mathbf{s} < \mathbf{x}_2'\mathbf{t})] + o(1) \\
&= \mathbb{E}[f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)|\mathbf{x}_2'\mathbf{s} - \mathbf{x}_2'\mathbf{t}|] + o(1).
\end{aligned}$$

Since the remainder term is uniformly small in $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0^{\delta_n}$, the verification of Condition CRA$_0$(iv) is complete.

*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 above calculation for $\mathcal{L}_0$.

### A.7.5 Proof of Lemma MS

#### A.7.5.1 Consistency

For consistency, we use Conditions MS and K(i)-(iii), $\mathbb{E}\|\mathbf{x}_2\|^6 < \infty$, $h_n \to 0$, and $nh_n^3 \to \infty$. Recall that $\tilde{\mathbf{V}}_n^{\mathtt{MS}} = \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n)$, set $\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) = \mathbb{E}[\tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta})]$, and note that

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}} = \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) + [\bar{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) - \bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)] + [\tilde{\mathbf{V}}_n^{\mathtt{MS}} - \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) - \bar{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) + \bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)].$$

Consistency of $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ then follows from the following four results:

$$\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) = \mathbf{V}_0^{\mathtt{MS}} + o(1), \tag{A.6}$$

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) - \bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) = o_{\mathbb{P}}(1), \tag{A.7}$$

$$\bar{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) - \bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) = o_{\mathbb{P}}(1), \tag{A.8}$$

$$\tilde{\mathbf{V}}_n^{\mathtt{MS}} - \tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) - \bar{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}) + \bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0) = O_{\mathbb{P}}(n^{-5/6}h_n^{-5/2}). \tag{A.9}$$

Equation (A.6) holds because

$$\begin{aligned}
&- \mathbb{E}\left[\tilde{V}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)\right] \\
&= h_n^{-2}\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2'(2y-1)\dot{K}\left(\mathbf{x}'\boldsymbol{\beta}_0/h_n\right)\right] \\
&= h_n^{-1}\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2'\int\left[1 - 2F_{u|x_1,\mathbf{x}_2}(-vh_n|vh_n - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)\right]\dot{K}(v)f_{x_1|\mathbf{x}_2}(vh_n - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dv\right] \\
&= -2\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2'\int F_{u|x_1,\mathbf{x}_2}^{(1)}(-\xi|\xi - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)v\dot{K}(v)f_{x_1|\mathbf{x}_2}(vh_n - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dv\right]
\end{aligned}$$

where $\xi$ is a mean value between $vh_n$ and 0. By boundedness of $f_{x_1|\mathbf{x}_2}(v|\mathbf{x}_2)$ and $F_{u|x_1,\mathbf{x}_2}^{(1)}(-v|v - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)$ and continuity with respect to $v$, the dominated convergence theorem implies

$$\begin{aligned}
&\int \mathbf{x}_2\mathbf{x}_2'\int 2F_{u|x_1,\mathbf{x}_2}^{(1)}(-\xi|\xi - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)v\dot{K}(v)f_{x_1|\mathbf{x}_2}(vh_n - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2) \\
&\to 2\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2'F_{u|x_1,\mathbf{x}_2}^{(1)}(0| - \mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right]\int v\dot{K}(v)du = \mathbf{V}_0^{\mathtt{MS}}
\end{aligned}$$

where we use $\int v\dot{K}(v)dv = -1$.

Equation (A.7) follows directly because the variance of $\mathbb{V}[\tilde{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)] = O(n^{-1}h_n^{-3})$, as shown below.

For Equation (A.8),

$$
\begin{aligned}
&\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) \\
&= h_n^{-2}\int \mathbf{x}_2\mathbf{x}_2' \int \{1-2F_{u|x_1,\mathbf{x}_2}(-\mathbf{x}'\boldsymbol{\beta}_0|x_1,\mathbf{x}_2)\}\dot{K}(\{x_1+\mathbf{x}_2'\boldsymbol{\theta}\}/h_n) f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)dx_1 dF_{\mathbf{x}_2}(\mathbf{x}_2) \\
&= h_n^{-1}\int \mathbf{x}_2\mathbf{x}_2' \int \{1-2F_{u|x_1,\mathbf{x}_2}(-vh_n|vh_n-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\}\dot{K}(v+\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)h_n^{-1}) \\
&\qquad\qquad\qquad\qquad\qquad \times f_{x_1|\mathbf{x}_2}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2) \\
&= -2\int \mathbf{x}_2\mathbf{x}_2' \int vF^{(1)}_{u|x_1,\mathbf{x}_2}(-\zeta|\zeta-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v+\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)h_n^{-1}) \\
&\qquad\qquad\qquad\qquad\qquad \times f_{x_1|\mathbf{x}_2}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2).
\end{aligned}
$$

where $\zeta$ lies between $vh_n$ and 0, and does not depend on $\boldsymbol{\theta}$. Then,

$$
\left\|\bar{\mathbf{V}}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta})-\bar{\mathbf{V}}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)\right\| \le C\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|h_n^{-1}\left(\int B(v)dv\right)\mathbb{E}\|\mathbf{x}_2\|^3
$$

where we use boundedness of $F^{(1)}_{u|\mathbf{x}}$, $f_{x_1|\mathbf{x}_2}$, and Assumption K(ii). Then, by $\|\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}}-\boldsymbol{\theta}_0\| = O_{\mathbb{P}}(n^{-1/3}) = o_{\mathbb{P}}(h_n)$, which establishes the result.

Finally, we verify Equation (A.9). Using Pollard (1989, Theorem 4.2), it suffices to show

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

By the assumed Lipschitz condition on $\dot{K}$,

$$
\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le Cn^{-1/3}}\left|\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)-\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\right| \le CB(\mathbf{x}'\boldsymbol{\beta}_0/h_n)\|\mathbf{x}_2\|h_n^{-1}n^{-1/3},
$$

and therefore

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le Cn^{-1/3}}\left|\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)-\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\right|^2\|\mathbf{x}_2\mathbf{x}_2'\|^2\right] \\
&\le Cn^{-2/3}h_n^{-2}\mathbb{E}\left[B(\mathbf{x}'\boldsymbol{\beta}_0/h_n)^2\|\mathbf{x}_2\|^6\right] \\
&\le Cn^{-2/3}h_n^{-1}\sup_{x_1,\mathbf{x}_2} f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)\left(\int B(v)^2dv\right)\mathbb{E}\|\mathbf{x}_2\|^6,
\end{aligned}
$$

which verifies the final result.

#### A.7.5.2 Approximate Mean Squared Error

First we show that $\|\bar{\mathbf{V}}_n^{\mathtt{MS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{MS}})-\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}_0)\| = o_{\mathbb{P}}(h_n^2)+O_{\mathbb{P}}(n^{-1/3})$ for Equation (A.8) above. Changing variables and using a Taylor expansion, we have

$$
\begin{aligned}
\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) &= -h_n^{-2}\mathbb{E}\left[\mathbf{x}_2\mathbf{x}_2'\int\{1-2F_{u|x_1,\mathbf{x}_2}(-\mathbf{x}'\boldsymbol{\beta}_0|x_1,\mathbf{x}_2)\}\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)f_{x_1|\mathbf{x}_2}(x_1|\mathbf{x}_2)dx_1\right]\\
&= 2h_n^{-1}\int\mathbf{x}_2\mathbf{x}_2'\int(vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0))F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+2h_n^{-1}\int\mathbf{x}_2\mathbf{x}_2'\int|vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|^2F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+h_n^{-1}\int\mathbf{x}_2\mathbf{x}_2'\int\{vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}^3F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}^{(2)}(\xi_{1,n}-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+h_n^{-1}\int\mathbf{x}_2\mathbf{x}_2'\int|vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)|^2F_{u|x_1,\mathbf{x}_2}^{(2)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+h_n^{-1}\int\mathbf{x}_2\mathbf{x}_2'\int\{vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}^3F_{u|x_1,\mathbf{x}_2}^{(2)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}^{(1)}(\xi_{2,n}-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+\frac{h_n^{-1}}{3}\int\mathbf{x}_2\mathbf{x}_2'\int\{vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\}^3F_{u|x_1,\mathbf{x}_2}^{(3)}(-\xi_{3,n}|\xi_{3,n}-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\dot{K}(v)\\
&\qquad\qquad\times f_{x_1|\mathbf{x}_2}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)
\end{aligned}
$$

where $\xi_1$, $\xi_2$ and $\xi_3$ lie between $vh_n-\mathbf{x}_2'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$ and 0.

Using $\int v\dot{K}(v)dv=1$, $\int\dot{K}(v)dv=0$, and $\int v^2\dot{K}(v)dv=0$,

$$
\begin{aligned}
\bar{\mathbf{V}}_n^{\mathtt{MS}}(\boldsymbol{\theta}) &= 2\int F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'dF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad-4(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\int\mathbf{x}_2F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'dF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+h_n^2\int\int v^3\dot{K}(v)F_{u|x_1,\mathbf{x}_2}^{(1)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}^{(2)}(\xi_1-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad-2(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\int\mathbf{x}_2F_{u|x_1,\mathbf{x}_2}^{(2)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'dF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+h_n^2\int\mathbf{x}_2\mathbf{x}_2'\int v^3\dot{K}(v)F_{u|x_1,\mathbf{x}_2}^{(2)}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}^{(2)}(\xi_2-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+\frac{h_n^2}{3}\int\mathbf{x}_2\mathbf{x}_2'\int v^3\dot{K}(v)F_{u|x_1,\mathbf{x}_2}^{(3)}(-\xi_3|\xi_3-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\
&\quad+O(h_n\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2+h_n^{-1}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^3)
\end{aligned}
$$

where we use boundedness and continuity of derivatives of $F_{u|x_1,\mathbf{x}_2}$ and $f_{x_1|\mathbf{x}_2}$ and $\mathbb{E}\|\mathbf{x}_2\|^6 < \infty$.

For $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ as $n \to \infty$, we have

$$\begin{aligned}\lim_{n\to\infty} &\int \mathbf{x}_2\mathbf{x}_2' \int v^3 \dot{K}(v) F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(2)}_{x_1|\mathbf{x}_2}(\xi_1 - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) dv dF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &= \int v^3\dot{K}(v)dv \int F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' dF_{\mathbf{x}_2}(\mathbf{x}_2),\end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty} &\int \mathbf{x}_2\mathbf{x}_2' \int v^3 \dot{K}(v) F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(2)}_{x_1|\mathbf{x}_2}(\xi_2 - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) dv dF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &= \int v^3\dot{K}(v)dv \int F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' dF_{\mathbf{x}_2}(\mathbf{x}_2),\end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty} &\int \mathbf{x}_2\mathbf{x}_2' \int v^3 \dot{K}(v) F^{(3)}_{u|x_1,\mathbf{x}_2}(-\xi_3|\xi_3-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f_{x_1|\mathbf{x}_2}(vh_n - \mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2) dv dF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &= \int v^3\dot{K}(v)dv \int F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' dF_{\mathbf{x}_2}(\mathbf{x}_2).\end{aligned}$$

Because $\mathbb{P}[|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \le Cn^{-1/3}] \to 1$, we have $\|\bar{\mathbf{V}}^{\mathtt{MS}}_n(\hat{\boldsymbol{\theta}}_n) - \bar{\mathbf{V}}^{\mathtt{MS}}_n(\boldsymbol{\theta}_0)\| = o_{\mathbb{P}}(h_n^2) + O_{\mathbb{P}}(n^{-1/3})$. Note that the term $O_{\mathbb{P}}(n^{-1/3})$ does not depend on $h_n$.

Next, we derive the leading bias and variance of $\tilde{\mathbf{V}}^{\mathtt{MS}}_n(\boldsymbol{\theta}_0)$. Specifically, we show that

$$\bar{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0) = V^{\mathtt{MS}}_{0,kl} + h_n^2\left(-\mathsf{B}_{kl} + o(1)\right)$$

and

$$\mathbb{V}[\tilde{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)] = \frac{1}{nh_n^3}\left(\mathsf{V}_{kl} + o(1)\right)$$

where $\mathsf{B}_{kl}$ and $\mathsf{V}_{kl}$, $1 \le l,k \le d$, are defined in the statement of Lemma MS.

For the bias, a Taylor expansion gives

$$\begin{aligned}&\bar{\mathbf{V}}^{\mathtt{MS}}_n(\boldsymbol{\theta}_0) - \mathbf{V}^{\mathtt{MS}}_0\\ &= 2h_n \int F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' f^{(1)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^2\dot{K}(v)dv\\ &\quad + h_n^2 \int \mathbf{x}_2\mathbf{x}_2' F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\int v^3\dot{K}(v) f^{(2)}_{x_1|\mathbf{x}_2}(\xi_1 - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\quad + h_n \int F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2' f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^2\dot{K}(v)dv\\ &\quad + h_n^2 \int \mathbf{x}_2\mathbf{x}_2' F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\int f^{(1)}_{x_1|\mathbf{x}_2}(\xi_2 - \mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)v^3\dot{K}(v)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\quad + \frac{h_n^2}{3}\int \mathbf{x}_2\mathbf{x}_2' \int F^{(3)}_{u|x_1,\mathbf{x}_2}(-\xi_3|\xi_3 - \mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f_{x_1|\mathbf{x}_2}(vh_n|\mathbf{x}_2)v^3\dot{K}(v)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\end{aligned}$$

where $\xi_1$, $\xi_2$ and $\xi_3$ denote mean-values between 0 and $vh_n$. Boundedness of $f^{(s)}_{x_1|\mathbf{x}_2}(v|\mathbf{x}_2)$ and

$F^{(s)}_{u|x_1,\mathbf{x}_2}(-v|v+x_1,\mathbf{x}_2)$, $k=1,2,3$, and continuity of these functions with respect to $v$ imply that

$$\begin{aligned}&\int \mathbf{x}_2\mathbf{x}_2' F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\int v^3\dot{K}(v)f^{(2)}_{x_1|\mathbf{x}_2}(\xi_1-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\to \int \mathbf{x}_2\mathbf{x}_2' F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^3\dot{K}(v)dv,\end{aligned}$$

$$\begin{aligned}&\int \mathbf{x}_2\mathbf{x}_2' F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)\int f^{(1)}_{x_1|\mathbf{x}_2}(\xi_2-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)v^3\dot{K}(v)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\to \int \mathbf{x}_2\mathbf{x}_2' F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f^{(1)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^3\dot{K}(v)dv,\end{aligned}$$

$$\begin{aligned}&\int \mathbf{x}_2\mathbf{x}_2' \int F^{(3)}_{u|x_1,\mathbf{x}_2}(-\xi_3|\xi_3-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(vh_n|\mathbf{x}_2)v^3\dot{K}(v)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\to \int \mathbf{x}_2\mathbf{x}_2' F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^3\dot{K}(v)dv.\end{aligned}$$

Therefore, because $\int_{-\infty}^{\infty} v^2\dot{K}(v)dv=0$,

$$\bar{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)=V^{\mathtt{MS}}_{0,kl}+h_n^2\left(\mathsf{B}_{kl}+o(1)\right)$$

with $\mathsf{B}_{kl}$ the $(k,l)$-th element of

$$\begin{aligned}\mathsf{B}=\int \mathbf{x}_2\mathbf{x}_2' \Big[&F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\\ &+F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f^{(1)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\\ &+\frac{1}{3}F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\Big]dF_{\mathbf{x}_2}(\mathbf{x}_2)\int v^3\dot{K}(v)dv.\end{aligned}$$

Finally, the approximate variance is obtained by noting that

$$\mathbb{E}[\tilde{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)^2]=\frac{1}{nh_n^4}\mathbb{E}\left[\dot{K}\left(\frac{\mathbf{x}'\boldsymbol{\beta}_0}{h_n}\right)\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l\right]^2$$

and

$$\begin{aligned}\frac{1}{h_n}\mathbb{E}\left[\dot{K}\left(\frac{\mathbf{x}'\boldsymbol{\beta}_0}{h_n}\right)\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l\right]^2&=\int(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2\int\dot{K}(v)^2f_{x_1|\mathbf{x}_2}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dvdF_{\mathbf{x}_2}(\mathbf{x}_2)\\ &\to\int(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int\dot{K}(v)^2dv\end{aligned}$$

where the last line follows by dominated convergence theorem. Since $\mathbb{E}[\tilde{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)]=O(1)$,

$$\mathbb{V}[\tilde{V}^{\mathtt{MS}}_{n,kl}(\boldsymbol{\theta}_0)]=\frac{1}{nh_n^3}\left[\int(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)dF_{\mathbf{x}_2}(\mathbf{x}_2)\int\dot{K}(v)^2dv+o(1)\right].$$

### A.7.6 Proof of Corollary CMS

We verify that Condition CMS implies Condition CRA. Set $q_n = b_n^d$.

*Condition CRA(i).* For uniform manageability, as shown for maximum score estimator, the class $\{y\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ has a polynomial bound on the covering number uniform in discrete probability measures and $L_n(\cdot)$ does not depend on $\boldsymbol{\theta}$. See, e.g., van der Vaart and Wellner (1996) for details. We can take $\bar{m}_n(\mathbf{z}) = |L_n(\mathbf{w})|$ and $\mathbb{E}[L_n^2(\mathbf{w})] \leq Cb_n^{-d}$, which implies $q_n\mathbb{E}[\bar{m}_n(\mathbf{z})^2] = O(1)$.

For uniform convergence of $M_n(\boldsymbol{\theta})$ to $M_0(\boldsymbol{\theta})$,

$$
\begin{aligned}
M_n(\boldsymbol{\theta}) &= \int \mathbb{E}[\psi(x_1, \mathbf{x}_2, \mathbf{w})\{f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w})\}^{-1}\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\}|\mathbf{w} = \mathbf{v}b_n]L(\mathbf{v})f_{\mathbf{w}}(\mathbf{v}b_n)d\mathbf{v} \\
&= \int\int\int_{-\mathbf{x}_2'\boldsymbol{\theta}}^{\infty} \psi(x, \mathbf{x}_2, \mathbf{v}b_n)dxdF_{\mathbf{x}_2}(\mathbf{x}_2|\mathbf{v}b_n)f_{\mathbf{w}}(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v} \\
&= \int \varphi_1(\mathbf{v}b_n; \boldsymbol{\theta})L(\mathbf{v})d\mathbf{v}.
\end{aligned}
$$

Using $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\varphi_1(\mathbf{w}; \boldsymbol{\theta}) - \varphi_1(\mathbf{0}; \boldsymbol{\theta})| \leq C\|\mathbf{w}\|^\epsilon$ for $\mathbf{w}$ close to $\mathbf{0}$,

$$
\begin{aligned}
\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})| &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int |\varphi_1(\mathbf{v}b_n; \boldsymbol{\theta}) - \varphi_1(\mathbf{0}; \boldsymbol{\theta})||L(\mathbf{v})|d\mathbf{v} \\
&\leq Cb_n^\epsilon \int \|\mathbf{v}\|^\epsilon |L(\mathbf{v})|d\mathbf{v} = o(1).
\end{aligned}
$$

Well-separatedness follows from uniqueness of maximizer, compactness of the parameter space, and continuity of $M_0(\boldsymbol{\theta}) = \mathbb{E}[y\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\}|\mathbf{w} = \mathbf{0}]$. Uniqueness of maximizer follows from Honoré and Kyriazidou (2000, Lemmas 6 and 7), compactness of $\boldsymbol{\Theta}$ is assumed, and continuity follows from differentiability of $M_0(\cdot)$ shown below.

*Condition CRA(ii).* By definition,

$$
M_0(\boldsymbol{\theta}) = \mathbb{E}[\psi(x_1, \mathbf{x}_2, \mathbf{w})\{f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w})\}^{-1}\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\}|\mathbf{w} = \mathbf{0}]
$$

and

$$
M_n(\boldsymbol{\theta}) = \int \mathbb{E}[\psi(x_1, \mathbf{x}_2, \mathbf{w})\{f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w})\}^{-1}\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\}|\mathbf{w} = \mathbf{v}b_n]L(\mathbf{v})f_{\mathbf{w}}(\mathbf{v}b_n)d\mathbf{v}.
$$

Since $\mathbb{E}[\psi(x_1, \mathbf{x}_2, \mathbf{w})\{f_{x_1}(x_1|\mathbf{x}_2, \mathbf{w})\}^{-1}\mathbb{1}\{x_1 + \mathbf{x}_2'\boldsymbol{\theta} \geq 0\}|\mathbf{w}] = \int\int_{-\mathbf{x}_2'\boldsymbol{\theta}}^{\infty} \psi(x, \mathbf{x}_2, \mathbf{w})dxdF_{\mathbf{x}_2}(\mathbf{x}_2|\mathbf{w})$, we have

$$
\frac{\partial}{\partial\boldsymbol{\theta}} M_n(\boldsymbol{\theta}) = \int\int \mathbf{x}_2\psi(-\mathbf{x}_2'\boldsymbol{\theta}, \mathbf{x}_2, \mathbf{v}b_n)dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n)f_{\mathbf{w}}(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v},
$$

$$
\frac{\partial}{\partial\boldsymbol{\theta}} M_0(\boldsymbol{\theta}) = \int \mathbf{x}_2\psi(-\mathbf{x}_2'\boldsymbol{\theta}, \mathbf{x}_2, \mathbf{0})dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{0})f_{\mathbf{w}}(\mathbf{0}),
$$

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_n(\boldsymbol{\theta}) = -\int\int \mathbf{x}_2\mathbf{x}_2'\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta},\mathbf{x}_2,\mathbf{v}b_n)dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n)f_{\mathbf{w}}(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v},$$

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_0(\boldsymbol{\theta}) = -\int \mathbf{x}_2\mathbf{x}_2'\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta},\mathbf{x}_2,\mathbf{0})dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{0})f_{\mathbf{w}}(\mathbf{0}),$$

where $\psi^{(s)}(x_1,\mathbf{x}_2,\mathbf{w}) = \frac{d^s}{dx_1^s}\psi(x_1,\mathbf{x}_2,\mathbf{w})$, $s\in\mathbb{Z}$, and we use $|\psi(x_1,\mathbf{x}_2,\mathbf{w})+\psi^{(1)}(x_1,\mathbf{x}_2,\mathbf{w})|\leq C$ for all $(x_1,\mathbf{x}_2,\mathbf{w})\in\mathcal{S}$, $\int\|\mathbf{x}_2\|^2dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v})\leq C$ for all $\mathbf{v}$ with $\|\mathbf{v}\|\leq\eta$, $f_{\mathbf{w}}(\cdot)$ is bounded around $\mathbf{0}$, and $\int|L(\mathbf{v})|d\mathbf{v}<\infty$, to invoke dominated convergence theorem.

Because the function $\mathbf{V}(\mathbf{w};\boldsymbol{\theta}) = \int \mathbf{x}_2\mathbf{x}_2'\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta},\mathbf{x}_2,\mathbf{w})dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{w})f_{\mathbf{w}}(\mathbf{w})$ satisfies $\|\mathbf{V}(\mathbf{w};\boldsymbol{\theta})-\varphi(\mathbf{0};\boldsymbol{\theta})\|\leq C\|\mathbf{w}\|^{\epsilon}$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta}$,

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta}}\left|\frac{\partial[M_n(\boldsymbol{\theta})-M_0(\boldsymbol{\theta})]}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right| &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta}}\int\|\mathbf{V}(\mathbf{v}b_n;\boldsymbol{\theta})-\mathbf{V}(\mathbf{0};\boldsymbol{\theta})\||L(\mathbf{v})|d\mathbf{v}\\ &\leq Cb_n^{\epsilon}\int\|\mathbf{v}\|^{\epsilon}|L(\mathbf{v})|d\mathbf{v} = o(1).\end{aligned}$$

For $\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}$,

$$\frac{\partial}{\partial\boldsymbol{\theta}}M_n(\boldsymbol{\theta}_0) = \int\varphi_2(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v} = \varphi_2(\mathbf{0}) + b_n^P\sum_{|\boldsymbol{\ell}=P}\int\mathbf{v}^{\boldsymbol{\ell}}\partial^{\boldsymbol{\ell}}\varphi_2(\tilde{\mathbf{v}})L(\mathbf{v})d\mathbf{v}$$

where we use multi-index notation, $\int\mathbf{v}^{\boldsymbol{\ell}}L(\mathbf{v})d\mathbf{v} = 0$ for $0<|\boldsymbol{\ell}|<P$, and $\tilde{\mathbf{v}}$ lies between $\mathbf{v}b_n$ and $\mathbf{0}$. Honoré and Kyriazidou (2000) show that, for any $(d_0,d_3)\in\{0,1\}^2$, $\mathbb{E}[A(d_0,d_3)|\mathbf{X},\alpha,\mathbf{w}=\mathbf{0}]>\mathbb{E}[(B(d_0,d_3)|\mathbf{X},\alpha,\mathbf{w}=\mathbf{0}]$ if and only if $x_1+\mathbf{x}_2'\boldsymbol{\theta}_0>0$, which implies $\mathbb{E}[A(d_0,d_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{X},\alpha,\mathbf{w}=\mathbf{0}] = \mathbb{E}[B(d_0,d_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{X},\alpha,\mathbf{w}=\mathbf{0}]$. Then,

$$\begin{aligned}&\psi(-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{0})\\ &=\mathbb{E}[A(Y_0,Y_3)-B(Y_0,Y_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0}]f_{x_1}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2,\mathbf{0})\\ &=\sum_{d_0,d_3\in\{0,1\}}\Pr(Y_0=d_0,Y_3=d_3|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0})f_{x_1}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2,\mathbf{0})\\ &\qquad\times\left\{\mathbb{E}[A(d_0,d_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0}]-\mathbb{E}[B(d_0,d_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0}]\right\}\\ &=\sum_{d_0,d_3\in\{0,1\}}\Pr(Y_0=d_0,Y_3=d_3|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0})f_{x_1}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2,\mathbf{0})\\ &\qquad\times\mathbb{E}[\mathbb{E}[A(d_0,d_3)-B(d_0,d_3)|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{X},\mathbf{w}=\mathbf{0},\alpha]|x_1=-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w}=\mathbf{0}]\\ &=0\end{aligned}$$

which implies $\varphi_2(\mathbf{0})=\mathbf{0}$. Finally, $b_n^P = o((nb_n^d)^{-1/3})$ implies that $\sqrt[3]{nq_n}\|\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\| = o(1)$ holds.

Positive definiteness of $\mathbf{V}_0^{\texttt{CMS}}$ is directly imposed in Condition CMS.

*Condition CRA(iii).* Uniform manageability follows from the class $\{y\mathbb{1}\{x_1+\mathbf{x}_2'\boldsymbol{\theta}\geq 0\}:\boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ having a polynomial bound on the uniform covering number, since $L_n(\cdot)$ does not depend on $\boldsymbol{\theta}$.

See, e.g., van der Vaart and Wellner (1996) for details. For this example,

$$\bar{d}_n^{\delta}(\mathbf{z}) = \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta} \mathbb{1}\{\mathbf{x}_2'\boldsymbol{\theta}_0 < -x_1 \leq \mathbf{x}_2'\boldsymbol{\theta}\}|L_n(\mathbf{w})| + \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta} \mathbb{1}\{\mathbf{x}_2'\boldsymbol{\theta} < -x_1 \leq \mathbf{x}_2'\boldsymbol{\theta}_0\}|L_n(\mathbf{w})|.$$

Conditional on $(\mathbf{x}_2, \mathbf{w})$, $x_1$ has bounded Lebesgue density and for $b_n$ sufficiently small,

$$\mathbb{E}[\bar{d}_n^{\delta}(\mathbf{z})^2] \leq C\mathbb{E}[\|\mathbf{x}_2\|L_n(\mathbf{w})^2]\delta \leq C\delta b_n^{-d}$$

by change of variables. The constant $C$ does not depend on $\delta$, and therefore $q_n \sup_{0<\delta'\leq\delta} \mathbb{E}[\bar{d}_n^{\delta'}(\mathbf{z})^2/\delta'] = O(1)$ holds.

*Condition CRA(iv).* Note that $\bar{d}_n^{\delta}(\mathbf{z})^3 = d_n^{\delta}(\mathbf{z})|L_n(\mathbf{w})|^2$ and $\bar{d}_n^{\delta}(\mathbf{z})^3 = d_n^{\delta}(\mathbf{z})|L_n(\mathbf{w})|^3$. Doing a similar calculation as above, we have $q_n^2\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^3] = O(\delta_n) = o(1)$ and $q_n^3 r_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = O(r_n^{-1}\delta_n) = o(1)$.

As in the Maximum Score example, $\mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s},\mathbf{s}) + \mathcal{C}_0^{\mathtt{CMS}}(\mathbf{t},\mathbf{t}) - 2\mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s},\mathbf{t}) > 0$ follows from its representation

$$\mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s},\mathbf{t}) = \mathbb{E}[\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2, \mathbf{0})\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}|\mathbf{w} = \mathbf{0}]f_{\mathbf{w}}(\mathbf{0}).$$

Now we verify the last condition.

$$\begin{aligned}
&\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n\mathbf{s}) - m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\}\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n\mathbf{t}) - m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\} \\
&= \{\mathbb{1}(\mathbf{x}_2'(\boldsymbol{\theta} + \delta_n\mathbf{s}) \geq -x_1) - \mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta} \geq -x_1)\}\{\mathbb{1}(\mathbf{x}_2'(\boldsymbol{\theta} + \delta_n\mathbf{t}) \geq -x_1) - \mathbb{1}(\mathbf{x}_2'\boldsymbol{\theta} \geq -x_1)\}|L_n(\mathbf{w})|^2 \\
&= \mathbb{1}(\delta_n \min\{\mathbf{x}_2'\mathbf{s}, \mathbf{x}_2'\mathbf{t}\} \geq -x_1 - \mathbf{x}_2'\boldsymbol{\theta} > 0)\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = 1 = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}|L_n(\mathbf{w})|^2 \\
&\quad + \mathbb{1}(\delta_n \max\{\mathbf{x}_2'\mathbf{s}, \mathbf{x}_2'\mathbf{t}\} < -x_1 - \mathbf{x}_2'\boldsymbol{\theta} \leq 0)\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = -1 = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}|L_n(\mathbf{w})|^2 \\
&= \mathbb{1}(\delta_n \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} \geq -x_1 - \mathbf{x}_2'\boldsymbol{\theta} > 0)\mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}|L_n(\mathbf{w})|^2
\end{aligned}$$

and

$$\begin{aligned}
&q_n\delta_n^{-1}\mathbb{E}[\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n\mathbf{s}) - m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\}\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n\mathbf{t}) - m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\}] \\
&= q_n^{-1}\delta_n^{-1}\int\int\{F_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2, \mathbf{w}) - F_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta} - \delta_n\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\}|\mathbf{x}_2, \mathbf{w})\} \\
&\qquad\qquad \times \mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{w})f_{\mathbf{w}}(\mathbf{w})L(\mathbf{w}/b_n)d\mathbf{w} \\
&= \delta_n^{-1}\int\int\{F_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}|\mathbf{x}_2, \mathbf{v}b_n) - F_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta} - \delta_n\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\}|\mathbf{x}_2, \mathbf{v}b_n)\} \\
&\qquad\qquad \times \mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n)f_{\mathbf{w}}(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v} \\
&= \int\int \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\}f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta} + \tilde{\delta}_n|\mathbf{x}_2, \mathbf{v}b_n) \\
&\qquad\qquad \times \mathbb{1}\{\operatorname{sgn}(\mathbf{x}_2'\mathbf{s}) = \operatorname{sgn}(\mathbf{x}_2'\mathbf{t})\}dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n)f_{\mathbf{w}}(\mathbf{v}b_n)L(\mathbf{v})d\mathbf{v}
\end{aligned}$$

where $\tilde{\delta}_n$ lies between 0 and $-\delta_n \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\}$. Then,

$$\begin{aligned}
&\Big| q_n \delta_n^{-1} \mathbb{E}[\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n \mathbf{s}) - m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\}\{m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta} + \delta_n \mathbf{t}) m_n^{\mathtt{CMS}}(\mathbf{z}, \boldsymbol{\theta})\}] \\
&\qquad - \int\int \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2, \mathbf{v}b_n) \\
&\qquad\qquad \times \mathbb{1}\{\mathrm{sgn}(\mathbf{x}_2'\mathbf{s}) = \mathrm{sgn}(\mathbf{x}_2'\mathbf{t})\} dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n) f_{\mathbf{w}}(\mathbf{v}b_n) L(\mathbf{v}) d\mathbf{v} \Big| \\
&\leq C\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^{\epsilon} \sup_{\mathbf{w}\in\mathcal{W}} \mathbb{E}[\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\}^2 \|\mathbf{x}_2\| B_f(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)|\mathbf{w}] \int |L(\mathbf{v})| d\mathbf{v} = O(\delta_n^{\epsilon})
\end{aligned}$$

as $\boldsymbol{\theta} \in \boldsymbol{\theta}_0^{\delta_n}$. By dominated convergence theorem,

$$\begin{aligned}
&\lim_{n\to\infty} \int\int \min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2, \mathbf{v}b_n) \\
&\qquad\qquad \times \mathbb{1}\{\mathrm{sgn}(\mathbf{x}_2'\mathbf{s}) = \mathrm{sgn}(\mathbf{x}_2'\mathbf{t})\} dF_{\mathbf{x}_2|\mathbf{w}}(\mathbf{x}_2|\mathbf{v}b_n) f_{\mathbf{w}}(\mathbf{v}b_n) L(\mathbf{v}) d\mathbf{v} \Big| \\
&= \mathbb{E}[\min\{|\mathbf{x}_2'\mathbf{s}|, |\mathbf{x}_2'\mathbf{t}|\} f_{x_1|\mathbf{x}_2,\mathbf{w}}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2, \mathbf{0}) \mathbb{1}\{\mathrm{sgn}(\mathbf{x}_2'\mathbf{s}) = \mathrm{sgn}(\mathbf{x}_2'\mathbf{t})\}|\mathbf{w} = \mathbf{0}] f_{\mathbf{w}}(\mathbf{0}) \\
&= \mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s}, \mathbf{t})
\end{aligned}$$

and this convergence occurs uniformly on $\boldsymbol{\Theta}_0^{\delta_n}$ because it is independent of $\boldsymbol{\theta}$. Finally,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^{\delta_n}} |q_n \mathbb{E}[\{m_n(\mathbf{z}, \boldsymbol{\theta} + \delta_n \mathbf{s}) - m_n(\mathbf{z}, \boldsymbol{\theta})\}\{m_n(\mathbf{z}, \boldsymbol{\theta} + \delta_n \mathbf{t}) - m_n(\mathbf{z}, \boldsymbol{\theta})\}/\delta_n] - \mathcal{C}_0^{\mathtt{CMS}}(\mathbf{s}, \mathbf{t})| = o(1),$$

as desired.

*Condition CRA(v).* The first condition follows from $q_n \bar{d}_n^{\delta}(\mathbf{z}) \leq \sup_{\mathbf{v}} |L(\mathbf{v})|$.

The second condition follows from the calculation similar to the covariance kernel calculation.

### A.7.7 Proof of Lemma CMS

#### A.7.7.1 Consistency

Define $\bar{\mathbf{V}}_n^{\mathtt{CMS}}(\boldsymbol{\theta}) = \mathbb{E}[\tilde{\mathbf{V}}_n^{\mathtt{CMS}}(\boldsymbol{\theta})]$. It suffices to show $\|\bar{\mathbf{V}}_n^{\mathtt{CMS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}}) - \mathbf{V}_0^{\mathtt{CMS}}\| = o_{\mathbb{P}}(1)$ and $\|\tilde{\mathbf{V}}_n^{\mathtt{CMS}} - \bar{\mathbf{V}}_n^{\mathtt{CMS}}(\hat{\boldsymbol{\theta}}_n^{\mathtt{CMS}})\| = o_{\mathbb{P}}(1)$.

For the first requirement, changing variables, using a Taylor expansion, and using the fact that

$\int \dot{K}(v)dv = 0$, we have

$$
\begin{aligned}
-\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}) &= h_n^{-2}\mathbb{E}\left[y\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)\mathbf{x}_2\mathbf{x}_2'L_n(\mathbf{w})\right] \\
&= h_n^{-1}\int \mathbb{E}\left[\int \psi(vh_n-\mathbf{x}_2'\boldsymbol{\theta},\mathbf{x}_2,\mathbf{w})\dot{K}(v)dv\mathbf{x}_2\mathbf{x}_2'\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w} \\
&= \int \mathbb{E}\left[\int \psi^{(1)}(\vartheta-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})v\dot{K}(v)dv\mathbf{x}_2\mathbf{x}_2'\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w} \\
&\quad - h_n^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\int \mathbb{E}\left[\int \psi^{(1)}(\vartheta-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\dot{K}(v)dv\mathbf{x}_2\mathbf{x}_2\mathbf{x}_2'\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w},
\end{aligned}
$$

where $\vartheta$ lies in between 0 and $vh_n$. By boundedness of $\psi^{(1)}$, $\mathbb{E}[\|\mathbf{x}_2\|^3|\mathbf{w}] \le C$ for $\mathbf{w}\in\mathcal{W}$, and $\int|\dot{K}(v)|dv<\infty$, the last term is $o(1)$ for $\boldsymbol{\theta}=\boldsymbol{\theta}_0+o(h_n)$. By continuous differentiability of $\psi^{(1)}$ with respect to its first argument and boundedness of the derivative,

$$
\begin{aligned}
&\left|\int \mathbb{E}\left[\int \{\psi^{(1)}(\vartheta-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})-\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\}v\dot{K}(v)dv\mathbf{x}_2\mathbf{x}_2'\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}\right| \\
&\le C\int \mathbb{E}\left[\left(h_n\int v^2|\dot{K}(v)|dv+\|\mathbf{x}_2\|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\right)\|\mathbf{x}_2\|^2\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})|L_n(\mathbf{w})|d\mathbf{w}=o(1).
\end{aligned}
$$

Thus, for $\boldsymbol{\theta}=\boldsymbol{\theta}_0+o(h_n)$,

$$
\begin{aligned}
\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}) &= -\int v\dot{K}(v)dv\int \mathbb{E}[\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\mathbf{x}_2\mathbf{x}_2'|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}+o(1) \\
&= \mathbf{V}_0^{\texttt{CMS}}+o(1).
\end{aligned}
$$

Using $\|\hat{\boldsymbol{\theta}}_n^{\texttt{CMS}}-\boldsymbol{\theta}_0\|=O_{\mathbb{P}}(\sqrt[3]{nb_n^d})=o_{\mathbb{P}}(h_n)$, it follows that $\bar{\mathbf{V}}_n^{\texttt{CMS}}(\hat{\boldsymbol{\theta}}_n^{\texttt{CMS}})-\mathbf{V}_0^{\texttt{CMS}}=o_{\mathbb{P}}(1)$.

To verify $\tilde{\mathbf{V}}_n^{\texttt{CMS}}-\bar{\mathbf{V}}_n^{\texttt{CMS}}(\hat{\boldsymbol{\theta}}_n^{\texttt{CMS}})=o_{\mathbb{P}}(1)$, it suffices to show that, for any $\delta_n=O(r_n^{-1})$,

$$
\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta_n}\left\|\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta})-\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)-\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta})+\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)\right\|+\left\|\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)-\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)\right\|=o_{\mathbb{P}}(1).
$$

It can be shown that $\|\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)-\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)\|=o_{\mathbb{P}}(1)$ easily because below we calculate the convergence rate of the variance, that is, $\mathbb{V}[\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)]=O(n^{-1}h_n^{-3}b_n^{-d})$. For the rest, Pollard (1989, Theorem 4.2) implies that it is enough to verify

$$
n^{-1}h_n^{-4}\mathbb{E}\left[\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta_n}\left|\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)-\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\right|^2\|\mathbf{x}_2\|^4|L_n(\mathbf{w})|^2\right]=o(1).
$$

By the Lipschitz condition on $\dot{K}$,

$$
\left|\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)-\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\right|\le h_n^{-1}\|\mathbf{x}_2\|\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|B\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right).
$$

Then,

$$
\begin{aligned}
&n^{-1}h_n^{-4}\mathbb{E}\left[\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\le\delta_n}\left|\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}}{h_n}\right)-\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\right|^2\|\mathbf{x}_2\|^2|L_n(\mathbf{w})|^2\right]\\
&\le n^{-1}h_n^{-3}(h_n^{-1}\delta_n)^2\mathbb{E}\left[\int B(v)^2 f_{x_1}(vh_n|\mathbf{x}_2,\mathbf{w})dv\|\mathbf{x}_2\|^6|L_n(\mathbf{w})|^2\right]\\
&\le o(1)n^{-1}h_n^{-3}b_n^{-d}\int B(v)^2dv\int\mathbb{E}[\|\mathbf{x}_2\|^6|\mathbf{w}=\mathbf{v}b_n]L(\mathbf{v})d\mathbf{v}=o(1).
\end{aligned}
$$

#### A.7.7.2 Approximate Mean Squared Error

First we calculate the order of bias. Changing variables and using a Taylor approximation, we obtain:

$$
\begin{aligned}
\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)&=-h_n^{-2}\int\mathbb{E}\left[\psi(x_1,\mathbf{x}_2,\mathbf{w})\{f_{x_1}(x_1|\mathbf{x}_2,\mathbf{w})\}^{-1}\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)\mathbf{x}_2\mathbf{x}_2'\middle|\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}\\
&=-\int v\dot{K}(v)dv\int\mathbb{E}[\psi^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\mathbf{x}_2\mathbf{x}_2'|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}\\
&\quad-h_n\int v^2\dot{K}(v)dv\int\mathbb{E}[\psi^{(2)}(-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\mathbf{x}_2\mathbf{x}_2'|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}\\
&\quad-h_n^2\int\mathbb{E}[\int\psi^{(3)}(\vartheta-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2,\mathbf{w})\dot{K}(v)dv\mathbf{x}_2\mathbf{x}_2'|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})L_n(\mathbf{w})d\mathbf{w}
\end{aligned}
$$

where $\vartheta$ lies between 0 and $vh_n$. Using $\int v\dot{K}(v)dv=-1$, $\int v^2\dot{K}(v)=0$, and boundedness of $\psi^{(3)}$,

$$
\bar{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)=\int\mathbf{V}(\mathbf{w};\boldsymbol{\theta}_0)L_n(\mathbf{w})d\mathbf{w}+O(h_n^2)=\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}M_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)+O(h_n^2).
$$

Finally, for the variance, changing variables and using a Taylor approximation, we obtain:

$$
\begin{aligned}
\mathbb{V}[\tilde{V}_{n,kl}^{\texttt{CMS}}(\boldsymbol{\theta}_0)]&=\frac{1}{nh_n^4}\mathbb{E}\left[\dot{K}\left(\frac{x_1+\mathbf{x}_2'\boldsymbol{\theta}_0}{h_n}\right)^2(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2|L_n(\mathbf{w})|^2\right]-\frac{1}{n}[\bar{V}_{n,kl}^{\texttt{CMS}}(\boldsymbol{\theta}_0)]^2\\
&=\frac{1}{nh_n^3}\mathbb{E}\left[\int\dot{K}^2(v)f_{x_1}(vh_n-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2,\mathbf{w})dv(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l)^2|L_n(\mathbf{w})|^2\right]+O(n^{-1})\\
&\le\frac{C}{nh_n^3}\int\dot{K}^2(v)dv\int\mathbb{E}[\|\mathbf{x}_2\|^4|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})|L_n(\mathbf{w})|^2d\mathbf{w}+O(n^{-1})\\
&=O(n^{-1}h_n^{-3}b_n^{-d}),
\end{aligned}
$$

for all $(k,l)$ elements of $\tilde{\mathbf{V}}_n^{\texttt{CMS}}(\boldsymbol{\theta}_0)$.

### A.7.8 Proof of Theorem ID

Let

$$
\mathcal{Z}_n^*(x)=\hat{F}_n^*(x)-\hat{F}_n(x)+\hat{f}_n(x_0)(x-x_0)+\tilde{f}_n^{(1)}(x_0)(x-x_0)^2/2
$$

and define

$$\hat{s}_n^*(a) = \sup_{s \geq 0} \left\{ s : \mathcal{Z}_n^*(s) - as = \sup_{x \geq 0} \left[ \mathcal{Z}_n^*(x) - ax \right] \right\},$$

$$b = \sup_{x \geq 0} \left[ \mathcal{Z}_n^*(x) - ax \right].$$

Denote the LCM of a function $f$ restricted on $I$ by $L_I f$. Since linear functions are concave,

$$\begin{aligned} \hat{s}_n^*(a) < x &\Rightarrow \mathcal{Z}_n^*(x) - ax - b < 0 \\ &\Rightarrow L_{[0,\infty)} \mathcal{Z}_n^*(x) \leq ax + b. \end{aligned}$$

Now, for contradiction, suppose that the left-derivative of $L_{[0,\infty)} \mathcal{Z}_n^*(x)$ is strictly greater than $a$. Since $L_{[0,\infty)} \mathcal{Z}_n^*(x)$ is concave, its derivative, denoted by $DL_{[0,\infty)} \mathcal{Z}_n^*(\cdot)$, satisfies $DL_{[0,\infty)} \mathcal{Z}_n^*(\tilde{x}) > a$ for $\tilde{x} \leq x$. By definition, $L_{[0,\infty)} \mathcal{Z}_n^*(\hat{s}_n^*(a)) = a\hat{s}_n^*(a) + b$ and the derivative condition implies $L_{[0,\infty)} \mathcal{Z}_n^*(x) > ax + b$, which is a contradiction and thus proving $DL_{[0,\infty)} \mathcal{Z}_n^*(x) = \tilde{f}_n^*(x) \leq a$.

Now,

$$\begin{aligned} \tilde{f}_n^*(x) < a &\Rightarrow L_{[0,\infty)} \mathcal{Z}_n^*(s) < as + b \quad \text{for } s \geq x \\ &\Rightarrow \hat{s}_n^*(a) < x. \end{aligned}$$

For the first implication, by definition,

$$\lim_{\varepsilon \downarrow 0} \frac{L_{[0,\infty)} \mathcal{Z}_n^*(s - \varepsilon) - L_{[0,\infty)} \mathcal{Z}_n^*(s)}{-\varepsilon} = \tilde{f}_n^*(s) < a \quad \text{for } s \geq x.$$

Then, there exist $\eta > 0$ such that $a - \eta > \tilde{f}_n^*(s)$ and $\varepsilon_0 > 0$ such that $-[L_{[0,\infty)} \mathcal{Z}_n^*(s - \varepsilon) - L_{[0,\infty)} \mathcal{Z}_n^*(s)]/\varepsilon < a - \eta$ for $0 < \varepsilon < \varepsilon_0$. Then,

$$\begin{aligned} L_{[0,\infty)} \mathcal{Z}_n^*(s) &< \varepsilon(a - \eta) + L_{[0,\infty)} \mathcal{Z}_n^*(s - \varepsilon) \\ &\leq \varepsilon(a - \eta) + a(s - \varepsilon) + b \\ &= as - \varepsilon\eta + b \\ &< as + b. \end{aligned}$$

Therefore, we have

$$\hat{s}_n^*(a) < x \iff \tilde{f}_n^*(x) < a.$$

Then,

$$\mathbb{P}^* \left[ n^{1/3} [\tilde{f}_n^*(x_0) - \hat{f}_n(x_0)] < t \right] = \mathbb{P}^* \left[ \hat{s}_n^* \left( \hat{f}_n(x_0) + n^{-1/3} t \right) < x_0 \right].$$

Recall that $\hat{s}_n^*(a)$ is the maximizer of $\{\mathcal{Z}_n^*(s) - as : s \in [0,\infty)\}$.

$$\mathbb{P}^*\left[\hat{s}_n^*\left(\hat{f}_n(x_0) + n^{-1/3}t\right) < x_0\right] = \mathbb{P}^*\left[\operatorname*{argmax}_{x\in[0,\infty)}\left\{\mathcal{Z}_n^*(x) - [\hat{f}_n(x_0) + n^{-1/3}t]x\right\} < x_0\right]$$

where the argmax functional picks the largest value if the set of the maximizers consists of more than one element.

Letting $x = x_0 + n^{-1/3}s$ where $s$ is the free variable,

$$\begin{aligned}
&\operatorname*{argmax}_{x\in[0,\infty)}\left\{\mathcal{Z}_n^*(x) - [\hat{f}_n(x_0) + n^{-1/3}t]x\right\} \\
&= \operatorname*{argmax}_{x_0+n^{-1/3}s}\left\{\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - [\hat{f}_n(x_0) + n^{-1/3}t][x_0 + n^{-1/3}s]\right\} \\
&= x_0 + n^{-1/3}\operatorname*{argmax}_{s\in[-n^{1/3}x_0,\infty)}\left\{\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - [\hat{f}_n(x_0) + n^{-1/3}t][x_0 + n^{-1/3}s]\right\}
\end{aligned}$$

Shifting and scaling does not change the location of maximizer so

$$\begin{aligned}
&\operatorname*{argmax}_{s}\left\{\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - [\hat{f}_n(x_0) + n^{-1/3}t][x_0 + n^{-1/3}s]\right\} \\
&= \operatorname*{argmax}_{s}\left\{n^{2/3}\left[\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)n^{-1/3}s\right] - ts\right\}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{P}^*\left[\operatorname*{argmax}_{x\in[0,\infty)}\left\{\mathcal{Z}_n^*(x) - [\hat{f}_n(x_0) + n^{-1/3}t]x\right\} < x_0\right] \\
&= \mathbb{P}^*\left[\operatorname*{argmax}_{s\in[-n^{1/3}x_0,\infty)}\left\{n^{2/3}\left[\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)n^{-1/3}s\right] - ts\right\} < 0\right].
\end{aligned}$$

Now we want to establish weak convergence in probability of

$$\left\{n^{2/3}\left[\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)n^{-1/3}s\right] - ts : s \in \mathcal{S}\right\} \tag{A.10}$$

with any compact $\mathcal{S} \subset \mathbb{R}$.

Note that

$$\begin{aligned}
&n^{2/3}\left[\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)n^{-1/3}s\right] \\
&\quad = n^{2/3}\left[\left(\hat{F}_n^* - \hat{F}_n\right)(x_0 + n^{-1/3}s) - \left(\hat{F}_n^* - \hat{F}_n\right)(x_0)\right] + \frac{1}{2}\tilde{f}_n^{(1)}(x_0)s^2.
\end{aligned}$$

Theorem 3.2 (i) in Sen, Banerjee, and Woodroofe (2010) implies that

$$n^{2/3}\left[\left(\hat{F}_n^* - \hat{F}_n\right)(x_0 + n^{-1/3}s) - \left(\hat{F}_n^* - \hat{F}_n\right)(x_0)\right] \rightsquigarrow_{\mathbb{P}} \mathcal{W}(f(x_0)s)$$

as stochastic process indexed by $s \in \mathcal{S}$.

Then, the sequence of stochastic processes in (A.10) weakly converges in probability to

$$\mathcal{W}(f(x_0)s) + \frac{1}{2}f^{(1)}(x_0)s^2 - ts.$$

As noted in Problem 3.2.5 of van der Vaart and Wellner (1996),

$$\begin{aligned}&\operatorname*{argmax}_s \left\{\mathcal{W}(f(x_0)s) + \frac{1}{2}f^{(1)}(x_0)s^2 - ts\right\}\\ &\quad =_d |4f(x_0)\{f^{(1)}(x_0)\}^{-2}|^{1/3} \operatorname*{argmax}_s \left\{\mathcal{W}(s) - s^2\right\} - t/|f^{(1)}(x_0)|.\end{aligned}$$

Therefore, provided that

$$\hat{s}_n^* = \operatorname*{argmax}_s \left\{n^{2/3}\left[\mathcal{Z}_n^*(x_0 + n^{-1/3}s) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)n^{-1/3}s\right] - ts : s \in \mathcal{S}\right\}$$

is $O_{\mathbb{P}^*}(1)$, the argmax continuous mapping theorem implies

$$\mathbb{P}^*\left[n^{1/3}[\tilde{f}_n^*(x_0) - \hat{f}_n(x_0)] < t\right] \to_{\mathbb{P}} \mathbb{P}\left[|4f(x_0)f^{(1)}(x_0)|^{1/3} \operatorname*{argmax}_s \left\{\mathcal{W}(s) - s^2\right\} < t\right].$$

Now to show $\hat{s}_n^* = O_{\mathbb{P}^*}(1)$, define

$$\begin{aligned}\hat{g}_n^* &= \operatorname*{argmax}_s \left\{\mathcal{Z}_n^*(s + x_0) - \mathcal{Z}_n^*(x_0) - \hat{f}_n(x_0)s - tsn^{-1/3}\right\}\\ &= \operatorname*{argmax}_s \left\{\hat{F}_n^*(s + x_0) - \hat{F}_n(s + x_0) + \tilde{f}_n^{(1)}(x_0)s^2/2 - tsn^{-1/3}\right\}.\end{aligned}$$

By definition, $\hat{g}_n^* = n^{-1/3}\hat{s}_n^*$, so we want to show $n^{1/3}\hat{g}_n^* = O_{\mathbb{P}^*}(1)$.

Modifying the proofs of Lemmas A.6–A.8 with $\hat{M}_n^*(\boldsymbol{\theta}) = \hat{F}_n^*(\boldsymbol{\theta} + x_0)$, $\hat{M}_n(\boldsymbol{\theta}) = \hat{F}_n(\boldsymbol{\theta} + x_0)$, and $\tilde{M}_n^*(\boldsymbol{\theta}) = \hat{F}_n^*(\boldsymbol{\theta}+x_0) - \hat{F}_n(\boldsymbol{\theta}+x_0) + \tilde{f}_n^{(1)}(x_0)\boldsymbol{\theta}^2/2 - t\boldsymbol{\theta}n^{-1/3}$ (no need to center the objective function at $\hat{M}_n(\boldsymbol{\theta}_0)$), it can be shown that $n^{1/3}\hat{g}_n^* = O_{\mathbb{P}^*}(1)$ with probability approaching one.

## A.8 Rule-of-Thumb Bandwidth Selection

We provide details on the rule-of-thumb (ROT) bandwidth selection rules used in the simulations reported for the Maximum Score Estimation and Isotonic Density estimation examples.

### A.8.1 Maximum Score Estimation

To construct a ROT bandwidth in this example, we choose a reference model involving finite dimensional parameters and calculate/approximate the corresponding leading constants entering the approximate MSE of $\tilde{\mathbf{V}}_n^{\texttt{MS}}$ (example-specific plug-in estimate) for the case of Maximum Score estimation, and of $\tilde{\mathbf{V}}_n^{\texttt{ND}}$ (generic numerical derivative estimate).

Specifically, we assume $u|\mathbf{x} \sim \mathsf{Normal}(0, \sigma_u^2(\mathbf{x}))$ and $x_1|\mathbf{x}_2 \sim \mathsf{Normal}(\mu_1, \sigma_1^2)$, 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,

$$F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) = -\frac{1}{\sigma_u(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}\phi(0),$$

$$F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) = \frac{1}{\sigma_u^2(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}\phi^{(1)}(0) = 0,$$

$$\begin{aligned} F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) &= -\frac{1}{\sigma_u^3(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}\phi^{(2)}(0) \\ &\quad -\frac{\sigma_u^{(2)}(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)\sigma_u(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2) - 2\{\sigma_u^{(1)}(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)\}^2}{\sigma_u^3(-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)}\phi(0), \end{aligned}$$

where $\sigma_u^{(s)}(x_1, \cdot) = d^s\sigma_u(x_1, \cdot)/dx_1^s$, and

$$f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) = \frac{1}{\sigma_1}\phi\left(\frac{-\mathbf{x}_2'\boldsymbol{\theta}_0 - \mu_1}{\sigma_1}\right),$$

$$f^{(1)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) = \frac{1}{\sigma_1^2}\phi^{(1)}\left(\frac{-\mathbf{x}_2'\boldsymbol{\theta}_0 - \mu_1}{\sigma_1}\right) = \frac{\mathbf{x}_2'\boldsymbol{\theta}_0 + \mu_1}{\sigma_1^3}\phi\left(\frac{-\mathbf{x}_2'\boldsymbol{\theta}_0 - \mu_1}{\sigma_1}\right),$$

$$f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) = \frac{1}{\sigma_1^3}\phi^{(2)}\left(\frac{-\mathbf{x}_2'\boldsymbol{\theta}_0 - \mu_1}{\sigma_1}\right) = \left(\left(\frac{\mathbf{x}_2'\boldsymbol{\theta}_0 + \mu_1}{\sigma_1}\right)^2 - 1\right)\frac{1}{\sigma_1^3}\phi\left(\frac{-\mathbf{x}_2'\boldsymbol{\theta}_0 - \mu_1}{\sigma_1}\right).$$

#### A.8.1.1 Plug-in Estimator $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$

Given our reference model, we need to estimate

$$\begin{aligned} \mathsf{B}_{kl} &= -\mathbb{E}\Big[\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2'\mathbf{e}_l\Big(F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2) \\ &\quad + \frac{1}{3}F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0, \mathbf{x}_2)f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\Big)\Big]\mu_3(\dot{K}) \end{aligned}$$

where $\mu_p(K) = \int v^p K(u)dv$. Define

$$\begin{aligned} \hat{\mathsf{B}}_{kl} &= -\frac{1}{n}\sum_{i=1}^n \mathbf{e}_k'\mathbf{x}_{2i}\mathbf{x}_{2i}'\mathbf{e}_l\Big\{\hat{F}^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n, \mathbf{x}_{2i})\hat{f}^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i}) \\ &\quad + \frac{1}{3}\hat{F}^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n, \mathbf{x}_{2i})\hat{f}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i})\Big\}\mu_3(\dot{K}) \end{aligned}$$

where the preliminary estimators $\hat{\boldsymbol{\theta}}_n$, $\hat{F}^{(1)}_{u|x_1,\mathbf{x}_2}(\cdot)$, $\hat{F}^{(2)}_{u|x_1,\mathbf{x}_2}(\cdot)$, $\hat{f}_{x_1|\mathbf{x}_2}(\cdot)$ and $\hat{f}^{(2)}_{x_1|\mathbf{x}_2}(\cdot)$ are constructed using maximum likelihood for the parametric reference model (i.e., heteroskedastic Probit), together with a flexible parametric specification for $\sigma_u^2(\mathbf{x}) = \boldsymbol{\gamma}'\mathbf{p}(\mathbf{x})$ with $\mathbf{p}(\mathbf{x})$ denoting a polynomial expansion.

Similarly, the higher order variance is

$$\mathsf{V}_{kl} = \mathbb{E}\left[(\mathbf{e}_k'\mathbf{x}_2\mathbf{x}_2\mathbf{e}_l)^2 f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right]\int \dot{K}(v)^2 dv,$$

and, given our reference model, a ROT estimate is

$$\hat{\mathsf{V}}_{kl} = \int \dot{K}(v)^2 dv \frac{1}{n}\sum_{i=1}^{n}(\mathbf{e}_k'\mathbf{x}_{2i}\mathbf{x}_{2i}\mathbf{e}_l)^2 \hat{f}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i}).$$

#### A.8.1.2 Numerical Differentiation Estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$

The bias constant is

$$\mathsf{B}_{kl} = \frac{1}{6}\left(\frac{\partial^4 M_0(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_k^3\partial\boldsymbol{\theta}_l} + \frac{\partial^4 M_0(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_l^3}\right)$$

Given our reference model,

$$\begin{aligned}\frac{\partial^4 M_0(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_k^3\partial\boldsymbol{\theta}_l} &= -2\mathbb{E}\left[(\mathbf{e}_k'\mathbf{x}_2)^3\mathbf{e}_l'\mathbf{x}_2 F^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right] \\ &\quad + 6\mathbb{E}\left[(\mathbf{e}_k'\mathbf{x}_2)^3\mathbf{e}_l'\mathbf{x}_2 F^{(2)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(1)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right] \\ &\quad - 6\mathbb{E}\left[(\mathbf{e}_k'\mathbf{x}_2)^3\mathbf{e}_l'\mathbf{x}_2 F^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_2'\boldsymbol{\theta}_0,\mathbf{x}_2) f^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right],\end{aligned}$$

and similarly for $\partial^4 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_l^3$. Then, we can estimate the bias constant by

$$\begin{aligned}\hat{\mathsf{B}}_{kl} &= \frac{1}{n}\sum_{i=1}^{n}\left[(\mathbf{e}_k'\mathbf{x}_{2i})^3\mathbf{e}_l'\mathbf{x}_{2i} + (\mathbf{e}_l'\mathbf{x}_{2i})^3\mathbf{e}_k'\mathbf{x}_{2i}\right] \\ &\quad \times\left[\frac{1}{3}\hat{F}^{(3)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n,\mathbf{x}_{2i})\hat{f}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i}) + \hat{F}^{(1)}_{u|x_1,\mathbf{x}_2}(0|-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n,\mathbf{x}_{2i})\hat{f}^{(2)}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i})\right]\end{aligned}$$

where $\hat{F}^{(1)}_{\varepsilon|z,\mathbf{x}_2}$, $\hat{F}^{(3)}_{\varepsilon|z,\mathbf{x}_2}$, $\hat{f}_{z|\mathbf{x}_2}$, $\hat{f}^{(2)}_{z|\mathbf{x}_2}$, etc., are as discussed above.

Finally, the variance constant is

$$\mathsf{V}_{kl} = \frac{1}{8}\left[\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,\mathbf{e}_k+\mathbf{e}_l) + \mathcal{C}_0(\mathbf{e}_k-\mathbf{e}_l,\mathbf{e}_k-\mathbf{e}_l) - 2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,\mathbf{e}_k-\mathbf{e}_l) - 2\mathcal{C}_0(\mathbf{e}_k+\mathbf{e}_l,-\mathbf{e}_k+\mathbf{e}_l)\right]$$

$\mathcal{C}_0$ is defined in Corollary MS. Recall that $\mathcal{C}_0(\mathbf{s},\mathbf{t}) = [\mathcal{L}_0(\mathbf{s}) + \mathcal{L}_0(\mathbf{t}) - \mathcal{L}_0(\mathbf{s}-\mathbf{t})]/2$. Then,

$$\mathsf{V}_{kl} = \frac{1}{16}\left[2\mathcal{L}_0(\mathbf{e}_k) + 2\mathcal{L}_0(\mathbf{e}_l) - \mathcal{L}_0(\mathbf{e}_k+\mathbf{e}_l) - \mathcal{L}_0(-\mathbf{e}_k+\mathbf{e}_l)\right]$$

with $\mathcal{L}_0(\mathbf{s}) = \mathbb{E}\left[|\mathbf{x}_2'\mathbf{s}| f_{x_1|\mathbf{x}_2}(-\mathbf{x}_2'\boldsymbol{\theta}_0|\mathbf{x}_2)\right]$. Then, we can estimate the variance constant by

$$\hat{\mathsf{V}}_{kl} = \frac{1}{16}\frac{1}{n}\sum_{i=1}^{n}\left[2\left|\mathbf{x}_{2i}'\mathbf{e}_k\right| + 2\left|\mathbf{x}_{2i}'\mathbf{e}_l\right| - \left|\mathbf{x}_{2i}'(\mathbf{e}_k+\mathbf{e}_l)\right| - \left|\mathbf{x}_{2i}'(-\mathbf{e}_k+\mathbf{e}_l)\right|\right]\hat{f}_{x_1|\mathbf{x}_2}(-\mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n|\mathbf{x}_{2i}).$$

### A.8.2 Isotonic Density Estimation

Here we briefly outline the calculations leading to a ROT bandwidth selector for derivative of density estimators, which are used to construct an estimate of the drift in this example. The reference model is $z \sim \mathsf{Exponential}(\lambda)$ where $\lambda$ will be estimated from the data.

#### A.8.2.1 Plug-in Estimator $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$

The plug-in estimator is the usual kernel density estimator. Assuming we use the second-order kernel, the leading bias term is $h^2 f^{(3)}(x_0)/2 \int u^2 K(u)du$ and the leading variance term is $n^{-1}h^{-3}f(x_0)\int \dot{K}^2(u)du$. For the ROT bandwidth, we take $f(x) = \exp(-x/\lambda)/\lambda$ and $K(u) = \phi(u)$ and we use Integrated MSE (IMSE) as the criterion. That is, we minimize

$$h^4 \int \left\{ f^{(3)}(x) \right\}^2 dx \left( \int u^2 K(u)du \right)^2 /4 + \int \dot{K}^2(u)du/nh^3$$

with respect to $h$. Then, $\int u^2 K(u)du = 1$, $\int \dot{K}^2(u)du = \frac{1}{4\sqrt{\pi}}$, $\int \left\{ f^{(3)}(x) \right\}^2 dx = (2\lambda^7)^{-1}$. Thus, the IMSE equals

$$\frac{h^4}{8\lambda^7} + \frac{1}{4\sqrt{\pi}nh^3}$$

and the minimizer is

$$\hat{h}_{\mathtt{AMSE}} = \lambda \left( \frac{3}{2\sqrt{\pi}} \right)^{1/7} n^{-1/7}.$$

#### A.8.2.2 Numerical Differentiation Estimator $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$

The numerical differentiation estimator is

$$\frac{M_0(x_0+h) + M_0(x_0-h) - 2M_0(x_0)}{h^2},$$

and we take $M_0(x) = F(x)$. Then, the bias constant is

$$\frac{1}{12}\frac{d^4}{dx^4}F(x)|_{x=x_0} = \frac{f^{(3)}(x_0)}{12}.$$

The variance constant is $2f(x_0)$. As in the plug-in estimator case, we take the IMSE, which equals

$$\frac{h^4}{12^2} \int \left\{ f^{(3)}(x) \right\}^2 dx + \frac{2}{nh^3} = \frac{h^4}{288\lambda^7} + \frac{2}{nh^3}$$

and the ROT bandwidth is

$$\hat{h}_{\mathtt{AMSE}} = \lambda \, (432)^{1/7} \, n^{-1/7}.$$

## References


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

Groeneboom, P. (1985): "Estimating a Monotone Density," *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (L. M. Le Cam and R. A. Olshen, eds.)*, (2), 539–555.

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

Horowitz, J. L. (2002): "Bootstrap Critical Values for Tests Based on the Smoothed Maximum Score Estimator," *Journal of Econometrics*, 111(2), 141–167.

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

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of econometrics*, 3(3), 205–228.

——— (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of econometrics*, 27(3), 313–333.

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

Sen, B., M. Banerjee, and M. Woodroofe (2010): "Inconsistency of Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38(4), 1953–1977.

Seo, M. H., and T. Otsu (2018): "Local M-Estimation with Discontinuous Criterion for Dependent and Limited Observations," *Annals of Statistics*, 46(1), 344–369.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* Springer.

Wand, M., and M. Jones (1995): *Kernel Smoothing.* Chapman & Hall/CRC.

Table 1: Simulations, Maximum Score Estimator, 95% Confidence Intervals.

(a) $n = 1,000$, $S = 2,000$, $B = 2,000$

| | **DGP 1** | | | **DGP 2** | | | **DGP 3** | | |
|---|---|---|---|---|---|---|---|---|---|
| | $h, \epsilon$ | Coverage | Length | $h, \epsilon$ | Coverage | Length | $h, \epsilon$ | Coverage | Length |
| **Standard** | | | | | | | | | |
| | | 0.639 | 0.475 | | 0.645 | 0.480 | | 0.640 | 0.247 |
| **m-out-of-n** | | | | | | | | | |
| $m = \lceil n^{1/2} \rceil$ | | 0.998 | 1.702 | | 0.997 | 1.754 | | 0.999 | 1.900 |
| $m = \lceil n^{2/3} \rceil$ | | 0.979 | 1.189 | | 0.979 | 1.223 | | 0.985 | 0.728 |
| $m = \lceil n^{4/5} \rceil$ | | 0.902 | 0.824 | | 0.894 | 0.839 | | 0.904 | 0.447 |
| **Plug-in:** $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ | | | | | | | | | |
| $0.7 \cdot h_{\mathtt{MSE}}$ | 0.434 | 0.941 | 0.501 | 0.406 | 0.947 | 0.513 | 0.105 | 0.904 | 0.256 |
| $0.8 \cdot h_{\mathtt{MSE}}$ | 0.496 | 0.946 | 0.503 | 0.464 | 0.952 | 0.516 | 0.120 | 0.917 | 0.260 |
| $0.9 \cdot h_{\mathtt{MSE}}$ | 0.558 | 0.951 | 0.506 | 0.522 | 0.951 | 0.518 | 0.135 | 0.930 | 0.267 |
| $1.0 \cdot h_{\mathtt{MSE}}$ | 0.620 | 0.954 | 0.510 | 0.580 | 0.952 | 0.522 | 0.150 | 0.941 | 0.273 |
| $1.1 \cdot h_{\mathtt{MSE}}$ | 0.682 | 0.959 | 0.515 | 0.638 | 0.955 | 0.526 | 0.165 | 0.948 | 0.281 |
| $1.2 \cdot h_{\mathtt{MSE}}$ | 0.744 | 0.961 | 0.522 | 0.696 | 0.959 | 0.532 | 0.180 | 0.958 | 0.288 |
| $1.3 \cdot h_{\mathtt{MSE}}$ | 0.806 | 0.962 | 0.531 | 0.754 | 0.960 | 0.539 | 0.195 | 0.966 | 0.296 |
| $h_{\mathtt{AMSE}}$ | 0.385 | 0.938 | 0.499 | 0.367 | 0.941 | 0.510 | 0.119 | 0.917 | 0.260 |
| $\hat{h}_{\mathtt{AMSE}}$ | 0.446 | 0.947 | 0.509 | 0.415 | 0.949 | 0.518 | 0.155 | 0.941 | 0.275 |
| **Num Deriv:** $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ | | | | | | | | | |
| $0.7 \cdot \epsilon_{\mathtt{MSE}}$ | 0.980 | 0.912 | 0.431 | 0.904 | 0.891 | 0.422 | 0.203 | 0.864 | 0.216 |
| $0.8 \cdot \epsilon_{\mathtt{MSE}}$ | 1.120 | 0.922 | 0.442 | 1.033 | 0.897 | 0.432 | 0.232 | 0.888 | 0.228 |
| $0.9 \cdot \epsilon_{\mathtt{MSE}}$ | 1.260 | 0.929 | 0.460 | 1.163 | 0.909 | 0.448 | 0.261 | 0.904 | 0.238 |
| $1.0 \cdot \epsilon_{\mathtt{MSE}}$ | 1.400 | 0.939 | 0.484 | 1.292 | 0.919 | 0.469 | 0.290 | 0.917 | 0.248 |
| $1.1 \cdot \epsilon_{\mathtt{MSE}}$ | 1.540 | 0.943 | 0.514 | 1.421 | 0.928 | 0.497 | 0.319 | 0.928 | 0.257 |
| $1.2 \cdot \epsilon_{\mathtt{MSE}}$ | 1.680 | 0.948 | 0.549 | 1.550 | 0.932 | 0.531 | 0.348 | 0.939 | 0.265 |
| $1.3 \cdot \epsilon_{\mathtt{MSE}}$ | 1.820 | 0.955 | 0.590 | 1.679 | 0.935 | 0.568 | 0.377 | 0.947 | 0.274 |
| $\epsilon_{\mathtt{AMSE}}$ | 0.483 | 0.878 | 0.410 | 0.476 | 0.871 | 0.412 | 0.216 | 0.877 | 0.221 |
| $\hat{\epsilon}_{\mathtt{AMSE}}$ | 0.518 | 0.877 | 0.414 | 0.513 | 0.884 | 0.418 | 0.368 | 0.932 | 0.269 |

Notes:
(i) Panel **Standard** refers to standard nonparametric bootstrap, Panel **m-out-of-n** refers to $m$-out-of-$n$ nonparametric bootstrap with subsample $m$, Panel **Plug-in:** $\tilde{\mathbf{V}}_n^{\mathtt{MS}}$ refers to our proposed bootstrap-based implemented using the example-specific plug-in drift estimator, and Panel **Num Deriv:** $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ refers to our proposed bootstrap-based implemented using the generic numerical derivative drift estimator.
(ii) Column "$h$, $\epsilon$" reports tuning parameter value used or average across simulations when estimated, and Columns "Coverage" and "Length" report empirical coverage and average length of bootstrap-based 95% percentile confidence intervals, respectively.
(iii) $h_{\mathtt{MSE}}$ and $\epsilon_{\mathtt{MSE}}$ correspond to the simulation MSE-optimal choices, $h_{\mathtt{AMSE}}$ and $\epsilon_{\mathtt{AMSE}}$ correspond to the AMSE-optimal choices, and $\hat{h}_{\mathtt{AMSE}}$ and $\hat{\epsilon}_{\mathtt{AMSE}}$ correspond to the ROT feasible implementation of $\hat{h}_{\mathtt{AMSE}}$ and $\hat{\epsilon}_{\mathtt{AMSE}}$ described in the supplemental appendix.

Table 2: Simulations, Isotonic Density Estimator, 95% Confidence Intervals.

(a) $n = 1,000$, $S = 2,000$, $B = 2,000$

| | **DGP 1** | | | **DGP 2** | | | **DGP 3** | | |
|---|---|---|---|---|---|---|---|---|---|
| | $h, \epsilon$ | Coverage | Length | $h, \epsilon$ | Coverage | Length | $h, \epsilon$ | Coverage | Length |
| **Standard** | | | | | | | | | |
| | | 0.828 | 0.146 | | 0.808 | 0.172 | | 0.821 | 0.155 |
| **m-out-of-n** | | | | | | | | | |
| $m = \lceil n^{1/2} \rceil$ | | 1.000 | 0.438 | | 0.995 | 0.495 | | 0.998 | 0.452 |
| $m = \lceil n^{2/3} \rceil$ | | 0.989 | 0.314 | | 0.979 | 0.360 | | 0.989 | 0.328 |
| $m = \lceil n^{4/5} \rceil$ | | 0.953 | 0.235 | | 0.937 | 0.274 | | 0.948 | 0.248 |
| **Plug-in:** $\tilde{\mathbf{V}}_n^{\mathtt{ID}}$ | | | | | | | | | |
| $0.7 \cdot h_{\mathtt{MSE}}$ | 0.264 | 0.955 | 0.157 | 0.202 | 0.947 | 0.183 | 0.209 | 0.957 | 0.165 |
| $0.8 \cdot h_{\mathtt{MSE}}$ | 0.302 | 0.954 | 0.157 | 0.231 | 0.946 | 0.182 | 0.239 | 0.952 | 0.165 |
| $0.9 \cdot h_{\mathtt{MSE}}$ | 0.339 | 0.951 | 0.156 | 0.260 | 0.944 | 0.181 | 0.269 | 0.949 | 0.164 |
| $1.0 \cdot h_{\mathtt{MSE}}$ | 0.377 | 0.949 | 0.154 | 0.289 | 0.941 | 0.180 | 0.299 | 0.948 | 0.163 |
| $1.1 \cdot h_{\mathtt{MSE}}$ | 0.415 | 0.940 | 0.151 | 0.318 | 0.938 | 0.178 | 0.329 | 0.944 | 0.161 |
| $1.2 \cdot h_{\mathtt{MSE}}$ | 0.452 | 0.934 | 0.147 | 0.347 | 0.934 | 0.176 | 0.359 | 0.939 | 0.158 |
| $1.3 \cdot h_{\mathtt{MSE}}$ | 0.490 | 0.922 | 0.142 | 0.376 | 0.928 | 0.173 | 0.389 | 0.935 | 0.155 |
| | | | | | | | | | |
| $h_{\mathtt{AMSE}}$ | 0.380 | 0.949 | 0.154 | 0.300 | 0.940 | 0.180 | 0.333 | 0.943 | 0.161 |
| $\hat{h}_{\mathtt{AMSE}}$ | 0.364 | 0.950 | 0.155 | 0.290 | 0.941 | 0.180 | 0.401 | 0.930 | 0.154 |
| **Num Deriv:** $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ | | | | | | | | | |
| $0.7 \cdot \epsilon_{\mathtt{MSE}}$ | 0.726 | 0.954 | 0.158 | 0.527 | 0.947 | 0.183 | 0.554 | 0.952 | 0.165 |
| $0.8 \cdot \epsilon_{\mathtt{MSE}}$ | 0.830 | 0.956 | 0.159 | 0.602 | 0.947 | 0.182 | 0.633 | 0.950 | 0.164 |
| $0.9 \cdot \epsilon_{\mathtt{MSE}}$ | 0.933 | 0.956 | 0.160 | 0.678 | 0.944 | 0.181 | 0.712 | 0.949 | 0.163 |
| $1.0 \cdot \epsilon_{\mathtt{MSE}}$ | 1.037 | 0.956 | 0.159 | 0.753 | 0.942 | 0.180 | 0.791 | 0.948 | 0.162 |
| $1.1 \cdot \epsilon_{\mathtt{MSE}}$ | 1.141 | 0.955 | 0.159 | 0.828 | 0.940 | 0.179 | 0.870 | 0.946 | 0.161 |
| $1.2 \cdot \epsilon_{\mathtt{MSE}}$ | 1.244 | 0.956 | 0.160 | 0.904 | 0.936 | 0.177 | 0.949 | 0.943 | 0.160 |
| $1.3 \cdot \epsilon_{\mathtt{MSE}}$ | 1.348 | 0.960 | 0.163 | 0.979 | 0.935 | 0.176 | 1.028 | 0.940 | 0.159 |
| | | | | | | | | | |
| $\epsilon_{\mathtt{AMSE}}$ | 0.927 | 0.956 | 0.160 | 0.731 | 0.943 | 0.180 | 0.812 | 0.948 | 0.162 |
| $\hat{\epsilon}_{\mathtt{AMSE}}$ | 0.888 | 0.956 | 0.159 | 0.708 | 0.943 | 0.181 | 0.978 | 0.942 | 0.159 |

Notes:
(i) Panel **Standard** refers to standard nonparametric bootstrap, Panel **m-out-of-n** refers to $m$-out-of-$n$ nonparametric bootstrap with subsample $m$, Panel **Plug-in:** $\tilde{\mathbf{V}}_n^{\mathtt{ID}}$ refers to our proposed bootstrap-based implemented using the example-specific plug-in drift estimator, and Panel **Num Deriv:** $\tilde{\mathbf{V}}_n^{\mathtt{ND}}$ refers to our proposed bootstrap-based implemented using the generic numerical derivative drift estimator.
(ii) Column "$h$, $\epsilon$" reports tuning parameter value used or average across simulations when estimated, and Columns "Coverage" and "Length" report empirical coverage and average length of bootstrap-based 95% percentile confidence intervals, respectively.
(iii) $h_{\mathtt{MSE}}$ and $\epsilon_{\mathtt{MSE}}$ correspond to the simulation MSE-optimal choices, $h_{\mathtt{AMSE}}$ and $\epsilon_{\mathtt{AMSE}}$ correspond to the AMSE-optimal choices, and $\hat{h}_{\mathtt{AMSE}}$ and $\hat{\epsilon}_{\mathtt{AMSE}}$ correspond to the ROT feasible implementation of $\hat{h}_{\mathtt{AMSE}}$ and $\hat{\epsilon}_{\mathtt{AMSE}}$ described in the supplemental appendix.