Econometrica, Vol. 88, No. 5 (September, 2020), 2203–2219

BOOTSTRAP-BASED INFERENCE FOR CUBE ROOT ASYMPTOTICS

MATIAS D. CATTANEO
Department of Operations Research and Financial Engineering, Princeton University

MICHAEL JANSSON
Department of Economics, University of California at Berkeley and CREATES

KENICHI NAGASAWA
Department of Economics, University of Warwick

This paper proposes a valid bootstrap-based distributional approximation for M-estimators exhibiting a Chernoff (1964)-type limiting distribution. For estimators of this kind, the standard nonparametric bootstrap is inconsistent. The method proposed herein is based on the nonparametric bootstrap, but restores consistency 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. We illustrate the applicability of our results with four examples in econometrics and machine learning.

KEYWORDS: Cube root asymptotics, bootstrapping, maximum score, empirical risk minimization.

1. INTRODUCTION

IN A SEMINAL PAPER, Kim and Pollard (1990) studied 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 non-Gaussian limiting distribution is that of the maximizer of a Gaussian process. Kim and Pollard's results cover not only celebrated examples such as the maximum score estimator of Manski (1975) and the isotonic density estimator of Grenander (1956), but also more contemporary estimators arising in examples related to classification problems in machine learning (Mohammadi and van de Geer (2005)), nonparametric inference under shape restrictions (Groeneboom and Jongbloed (2018)), massive data M-estimation framework (Shi, Lu, and Song (2018)), and maximum score estimation in high-dimensional settings (Mukherjee, Banerjee, and Ritov (2019)). Moreover, Seo and Otsu (2018) recently generalized Kim and Pollard (1990) to allow for n-varying objective functions (n denotes the sample size), further widening the applicability of cube-root-type asymptotics. For example, their results cover the conditional maximum score estimator of Honoré and Kyriazidou (2000).

Matias D. Cattaneo: cattaneo@princeton.edu
Michael Jansson: mjansson@econ.berkeley.edu
Kenichi Nagasawa: Kenichi.Nagasawa@warwick.ac.uk

A previous version of this paper circulated under the title "Bootstrap-Based Inference for Cube Root Consistent Estimators." For comments and suggestions, we are grateful to Mehmet Caner, Andreas Hagemann, Kei Hirano, Bo Honoré, Guido Kuersteiner, Mykhaylo Shkolnikov, Ronnie Sircar, Ulrich Müller, Whitney Newey, and participants at various conferences, workshops, and seminars. Cattaneo gratefully acknowledges financial support from the National Science Foundation through Grants SES-1459931 and SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through Grants SES-1459967 and SES-1947662 and the research support of CREATES (funded by the Danish National Research Foundation under Grant DNRF78).

 https://doi.org/10.3982/ECTA17950

An important feature of Chernoff-type asymptotic distributional approximations is that the covariance kernel of the Gaussian process characterizing the limiting distribution often 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). 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. Unfortunately, however, the standard nonparametric bootstrap is well known to be invalid in this setting (Abrevaya and Huang (2005), Léger and MacGibbon (2006), Kosorok (2008), Sen, Banerjee, and Woodroofe (2010)). The purpose of this paper is to propose a generic and easy-to-implement bootstrap-based distributional approximation applicable in the context of cube root asymptotics.

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, our method offers a consistent distributional approximation for estimators exhibiting cube root asymptotics. 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.

This paper is not the first to propose a consistent resampling-based distributional approximation for cube-root-type estimators. For canonical 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). Related applicable methods are the m-out-of-n bootstrap (Bickel, Götze, and van Zwet (1997)), whose applicability was discussed and extended by Lee and Pun (2006) and Lee and Yang (2020), the rescaled bootstrap (Dümbgen (1993)), and the numerical bootstrap (Hong and Li (2020)). In addition, case-specific (smooth or non-standard) bootstrap methods have been proposed for leading examples such as monotone density estimation (Kosorok (2008), Sen, Banerjee, and Woodroofe (2010)), maximum score estimation (Patra, Seijo, and Sen (2018)), and the current status model (Groeneboom and Hendrickx (2018)). For the more generic cube-root-type estimators analyzed in Seo and Otsu (2018), subsampling appears to be the only method available, and indeed the authors discussed in their concluding remarks the need for (and importance of) developing resampling methods based on the standard nonparametric bootstrap. Our paper appears to be the first to provide one such method.

Like ours, each of the resampling methods mentioned above can be viewed as offering a "robust" alternative to the standard nonparametric bootstrap but, unlike ours, existing methods achieve consistency by modifying the distribution used to generate the bootstrap sample. In contrast, our bootstrap-based method achieves consistency by means of an analytic modification of the objective function used to construct the bootstrap-based distributional approximation. As further discussed below, this approach results in a procedure that is conceptually related to the bootstrap methods developed by Andrews and Soares (2010) and Fang and Santos (2019) in other econometrics contexts.

Implementation of our procedure is not computationally demanding. Indeed, the only ingredient needed to implement our modification on the objective function is a consistent estimator of a certain Hessian matrix. We propose a generic estimator based on numerical derivatives and present a consistency result as well as an approximate mean squared error expansion for that estimator. In addition, we illustrate how example-specific features can be sometimes exploited to construct alternative estimators.

The paper proceeds as follows. Section 2 is heuristic and outlines the main idea underlying our approach in the M-estimation setting of Kim and Pollard (1990). Section 3 then makes that heuristic discussion rigorous in a more general setting similar to that of Seo and Otsu (2018). Section 4 illustrates our bootstrap-based inference method with four examples: the maximum score estimator of Manski (1975, 1985), the conditional maximum score panel data estimator of Manski (1987), the conditional maximum score dynamic panel data estimator of Honoré and Kyriazidou (2000), and the classification estimator of Mohammadi and van de Geer (2005). Section 5 reports simulation evidence for the case of the maximum score estimator, and Section 6 concludes. Section 7 describes the proof of our main result, while the Supplemental Material (Cattaneo, Jansson, and Nagasawa (2020)) contains omitted proofs and details.

2. HEURISTICS

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

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

where m_0 is a known function, and where $\mathbf{z}$ is a random vector of which a random sample $\mathbf{z}_1, \dots, \mathbf{z}_n$ 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:

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

where $\rightsquigarrow$ denotes weak convergence, $\mathcal{G}_0$ is a non-degenerate zero-mean Gaussian process, and $\mathcal{Q}_0(\mathbf{s}) = -\mathbf{s}'\mathbf{H}_0\mathbf{s}/2$, where $\mathbf{H}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$.

Whereas the matrix $\mathbf{H}_0$ 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 $\mathbf{z}^*_{1,n}, \dots, \mathbf{z}^*_{n,n}$ denote a random sample from the empirical distribution of $\mathbf{z}_1, \dots \mathbf{z}_n$ 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 $\mathbb{P}[\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \le \cdot]$ is given by $\mathbb{P}_n^*[\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n \le \cdot]$, where $\mathbb{P}_n^*$ 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 (Abrevaya and Huang (2005), Léger and MacGibbon (2006), Kosorok (2008), Sen, Banerjee, and Woodroofe (2010)).

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

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

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

$$\hat{G}_n(\mathbf{s}) = n^{2/3}\big[\hat{M}_n(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) - \hat{M}_n(\boldsymbol{\theta}_0) - M_0(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) + M_0(\boldsymbol{\theta}_0)\big] \tag{4}$$

is a zero-mean random process, while

$$Q_n(\mathbf{s}) = n^{2/3}\big[M_0(\boldsymbol{\theta}_0 + \mathbf{s}n^{-1/3}) - M_0(\boldsymbol{\theta}_0)\big] \tag{5}$$

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, $\hat{G}_n$ and Q_n are usually asymptotically Gaussian and asymptotically quadratic, respectively, in the sense that

$$\hat{G}_n(\mathbf{s}) \rightsquigarrow \mathcal{G}_0(\mathbf{s}) \tag{6}$$

and

$$Q_n(\mathbf{s}) \to \mathcal{Q}_0(\mathbf{s}). \tag{7}$$

Under regularity conditions ensuring among other things that the convergence in (6) and (7) is suitably uniform in $\mathbf{s}$, (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

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

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

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

and

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

Under mild conditions, $\hat{G}_n^*$ satisfies the following bootstrap counterpart of (6):

$$\hat{G}_n^*(\mathbf{s}) \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0(\mathbf{s}), \tag{8}$$

where $\rightsquigarrow_{\mathbb{P}}$ denotes conditional weak convergence in probability (defined as in 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 $\hat{Q}_n(\mathbf{s}) \nrightarrow_{\mathbb{P}} \mathcal{Q}_0(\mathbf{s})$ in general. In other words, the natural bootstrap counterpart of (7) typically fails and, as a partial consequence, so does the natural bootstrap counterpart of (2); that is, $\sqrt[3]{n}(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \not\rightsquigarrow_{\mathbb{P}} \operatorname{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{G}_0(\mathbf{s}) + \mathcal{Q}_0(\mathbf{s})\}$.

To the extent that the 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, $\tilde{G}_n^*$ should be asymptotically equivalent to $\hat{G}_n^*$, where $\tilde{G}_n^*$ is the counterpart of $\hat{G}_n^*$ associated with $\tilde{m}_n$:

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

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

Second, and most importantly, $\tilde{Q}_n$ should be asymptotically quadratic, where $\tilde{Q}_n$ is the counterpart of $\hat{Q}_n$ associated with $\tilde{m}_n$:

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

Accordingly, let

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

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

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

where, by construction, $\tilde{G}_n^*(\mathbf{s}) = \hat{G}_n^*(\mathbf{s})$ and $\tilde{Q}_n(\mathbf{s}) = -\mathbf{s}'\tilde{\mathbf{H}}_n\mathbf{s}/2$. Because $\tilde{G}_n^* = \hat{G}_n^*$, $\tilde{G}_n^*(\mathbf{s}) \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0(\mathbf{s})$ whenever (8) holds. In addition, we have $\tilde{Q}_n(\mathbf{s}) \to_{\mathbb{P}} \mathcal{Q}_0(\mathbf{s})$ provided $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}_0$. As a consequence, it seems plausible that if $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}_0$, then our proposed bootstrap-based distributional approximation will be valid in the sense that $\sqrt[3]{n}(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \rightsquigarrow_{\mathbb{P}} \operatorname{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{G}_0(\mathbf{s}) + \mathcal{Q}_0(\mathbf{s})\}$.

For the purposes of situating this paper in the literature, the following alternative heuristic explanation of our approach may be useful. Restating the result in (2) as

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

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

$\tilde{\mathcal{S}}_n(\mathcal{G}) = \operatorname{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{G}(\mathbf{s}) + \tilde{\mathcal{Q}}_n(\mathbf{s})\}$ is an estimator of $\mathcal{S}_0(\mathcal{G})$. In other words, our procedure replaces the functional $\mathcal{S}_0$ with a consistent estimator (namely, $\tilde{\mathcal{S}}_n$) and its random argument $\mathcal{G}_0$ with a bootstrap approximation (namely, $\hat{G}_n^*$). The same type of generic construction has appeared in the econometrics literature before, notably in Andrews and Soares (2010) and Fang and Santos (2019).

Our bootstrap-based distributional approximation can be shown to be 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 (first-order) 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.

3. MAIN RESULT

When making the heuristics of Section 2 precise, we 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

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

In other words, in the setting considered in this section, $\hat{\boldsymbol{\theta}}_n$ 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. This generalization 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 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

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

and

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

$$\boldsymbol{\Theta}_0^{\delta} = \big\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta\big\}.$$

Condition CRA (Cube Root Asymptotics) For some $q_n > 0$ with $r_n = \sqrt[3]{nq_n} \to \infty$, the following are satisfied:

(i) The class $\{\mathcal{M}_n : n \geq 1\}$ 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 $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta} \|\partial^2[M_n(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\| = o(1)$. Also, $r_n\|\partial M_n(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\| = o(1)$ and $\mathbf{H}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(iii) For some $\delta > 0$, the class $\{\mathcal{D}_n^{\delta'} : n \geq 1, 0 < \delta' \leq \delta\}$ 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 $\delta_n > 0$ with $\delta_n = O(r_n^{-1})$, $q_n^3 r_n^{-1}\mathbb{E}[\bar{d}_n^{\delta_n}(\mathbf{z})^4] = o(1)$ 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 $\mathbf{s} \neq \mathbf{t}$,

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

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

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

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 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 differences between their assumptions and Condition CRA are technical in nature, since we need to slightly strengthen their assumptions in order to be able to analyze the bootstrap. For instance, the displayed part of Condition CRA(iv) is a locally uniform (with respect to $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$) version of its counterpart in Kim and Pollard (1990). 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) are also technical in nature, but for completeness we highlight two here. First, they controlled the complexity of various function classes using the concept of bracketing entropy, while we 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) controlled the bias of $\hat{\boldsymbol{\theta}}_n$ through a locally uniform bound on $M_n - M_0$, Condition CRA controls the bias through the first and second derivatives of $M_n - M_0$.

Under Condition CRA, the effective sample size is $nq_n = r_n^3$ and if $\hat{\boldsymbol{\theta}}_n$ is an approximate maximizer of $\hat{M}_n$, then $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ has a limiting distribution of Chernoff (1964) type. The heuristics of the previous section are rate-adaptive (i.e., $\sqrt[3]{n}$ can be replaced by a generic r_n), so once again it stands to reason that if $\tilde{\mathbf{H}}_n$ is a consistent estimator of $\mathbf{H}_0$, then the distribution of $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ can be consistently estimated by that of $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)$, where $\tilde{\boldsymbol{\theta}}_n^*$

is an approximate maximizer of

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

with $\mathbf{z}_{1,n}^*, \dots, \mathbf{z}_{n,n}^*$ being a random sample from the empirical distribution of $\mathbf{z}_1, \dots, \mathbf{z}_n$. A precise statement is given in the following theorem.

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

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

then

$$r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \rightsquigarrow \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{G}_0(\mathbf{s}) + \mathcal{Q}_0(\mathbf{s})\}, \tag{9}$$

and

$$r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) \rightsquigarrow_{\mathbb{P}} \operatorname*{argmax}_{\mathbf{s}\in\mathbb{R}^d}\{\mathcal{G}_0(\mathbf{s}) + \mathcal{Q}_0(\mathbf{s})\}, \tag{10}$$

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

The algorithm for our bootstrap-based distributional approximation is as follows:
Step 1. Using the sample $\mathbf{z}_1, \dots, \mathbf{z}_n$, compute $\hat{\boldsymbol{\theta}}_n$ by approximately maximizing $\hat{M}_n(\boldsymbol{\theta})$.
Step 2. Using $\hat{\boldsymbol{\theta}}_n$ and $\mathbf{z}_1, \dots, \mathbf{z}_n$, compute $\tilde{\mathbf{H}}_n$. (See Section 3.1 for a generic estimator $\tilde{\mathbf{H}}_n$.)
Step 3. Using $\hat{\boldsymbol{\theta}}_n$, $\tilde{\mathbf{H}}_n$, and the bootstrap sample $\mathbf{z}_{1,n}^*, \dots, \mathbf{z}_{n,n}^*$, compute $\tilde{\boldsymbol{\theta}}_n^*$ by approximately maximizing $\tilde{M}_n^*(\boldsymbol{\theta})$. ($\hat{\boldsymbol{\theta}}_n$ and $\tilde{\mathbf{H}}_n$ are not recomputed at this step.)
Step 4. Repeat *Step 3* to generate draws from the distribution of $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n)$.

3.1. *Estimation of* $\mathbf{H}_0$

A generic numerical derivative estimator of $\mathbf{H}_0$ is the matrix $\tilde{\mathbf{H}}_n^{\text{ND}}$ with element (k, l) given by

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

where $\mathbf{e}_k$ is the kth unit vector in $\mathbb{R}^d$ and where ϵ_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* $\tilde{\mathbf{H}}_n^{\text{ND}} \to_{\mathbb{P}} \mathbf{H}_0$.

Plausibility of the high-level condition $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ follows from (9). To facilitate practical implementation, it is useful to go beyond consistency and develop a Nagar-type mean squared error (MSE) expansion that can be used to select ϵ_n. To state one such result for $\tilde{H}^{\mathrm{ND}}_{n,kl}$, define

$$\ddot{M}_{n,kl}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_k\,\partial\theta_l} M_n(\boldsymbol{\theta}), \qquad \ddot{M}_{0,kl}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_k\,\partial\theta_l} M_0(\boldsymbol{\theta}),$$
$$\mathsf{B}_{kl} = -\frac{1}{6}\left[\frac{\partial^2}{\partial\theta_k^2}\ddot{M}_{0,kl}(\boldsymbol{\theta}_0) + \frac{\partial^2}{\partial\theta_l^2}\ddot{M}_{0,kl}(\boldsymbol{\theta}_0)\right],$$

and

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

LEMMA 2: *Suppose the conditions of Lemma 1 hold and that, for some* $\delta > 0$, $\ddot{M}_{0,kl}$ *and* $\ddot{M}_{n,kl}$ *are twice continuously differentiable on* $\boldsymbol{\Theta}_0^\delta$ *with* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0^\delta}\|\partial^2[\ddot{M}_{n,kl}(\boldsymbol{\theta}) - \ddot{M}_{0,kl}(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\| = o(1)$. *If* $\mathcal{C}_0(\mathbf{s}, -\mathbf{s}) = 0$ *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* $\tilde{H}^{\mathrm{ND}}_{n,kl}$ *admits an approximation* $\check{H}^{\mathrm{ND}}_{n,kl}$ *satisfying*

$$\tilde{H}^{\mathrm{ND}}_{n,kl} = \check{H}^{\mathrm{ND}}_{n,kl} + o_{\mathbb{P}}\left(\epsilon_n^2 + \frac{1}{\sqrt{r_n^3\epsilon_n^3}}\right) + O_{\mathbb{P}}\left(\frac{1}{r_n}\right),$$

where the $O_{\mathbb{P}}(1/r_n)$ *term does not depend on* ϵ_n *and where*

$$\mathbb{E}\big[\big(\check{H}^{\mathrm{ND}}_{n,kl} - H_{n,kl}\big)^2\big] = \epsilon_n^4\mathsf{B}_{kl}^2 + \frac{1}{r_n^3\epsilon_n^3}\mathsf{V}_{kl} + o\left(\epsilon_n^4 + \frac{1}{r_n^3\epsilon_n^3}\right), \qquad H_{n,kl} = -\ddot{M}_{n,kl}(\boldsymbol{\theta}_0).$$

The conditions $\mathcal{C}_0(\mathbf{s}, -\mathbf{s}) = 0$ and $\mathcal{C}_0(\mathbf{s}, \mathbf{t}) = \mathcal{C}_0(-\mathbf{s}, -\mathbf{t})$ are satisfied in all of the examples we have analyzed. Using the lemma, the approximate MSE (AMSE), $\epsilon_n^4\mathsf{B}_{kl}^2 + r_n^{-3}\epsilon_n^{-3}\mathsf{V}_{kl}$, can be minimized by choosing ϵ_n proportional to $r_n^{-3/7}$, the optimal factor of proportionality being a function of B_{kl} and V_{kl}. To be specific, the optimal ϵ_n is given by $\epsilon^{\mathtt{AMSE}}_{n,kl} = (3\mathsf{V}_{kl}/4\mathsf{B}_{kl}^2)^{1/7} r_n^{-3/7}$, a feasible version of which can be constructed by replacing B_{kl} and V_{kl} with preliminary estimators thereof.

4. EXAMPLES

4.1. *Maximum Score*

To describe a version of the maximum score estimator of Manski (1975, 1985), suppose $\mathbf{z}_1, \dots, \mathbf{z}_n$ is a random sample of $\mathbf{z} = (y, \mathbf{x}')'$ generated by the binary response model

$$y = \mathbb{1}\big(\mathbf{x}'\boldsymbol{\beta}_0 + u \geq 0\big), \qquad \mathrm{Median}(u|\mathbf{x}) = 0,$$

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 u is an unobserved error term. 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 $\boldsymbol{\theta}_0$ is any $\hat{\boldsymbol{\theta}}_n^{\mathrm{MS}}$ approximately maximizing $\hat{M}_n$ for

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

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. Second, in addition to the generic estimator $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ discussed above, the maximum score estimator admits example-specific consistent estimators of the $\mathbf{H}_0$ associated with it.

Under standard regularity conditions (stated in Section A.2 of the Supplemental Material), Condition CRA is satisfied with $q_n = 1$,

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

and

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

where $f_{a|\mathbf{b}}$ denotes the conditional Lebesgue density of a given $\mathbf{b}$. As a consequence, Theorem 1 is applicable to $\hat{\boldsymbol{\theta}}_n^{\mathrm{MS}}$ and the consistency requirement $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}^{\mathrm{MS}}$ is satisfied by the numerical derivative estimator discussed in Section 3.1 if $\epsilon_n \to 0$ and $n\epsilon_n^3 \to \infty$. Under the additional regularity conditions of Lemma 2, MSE-optimal tuning parameter choices are feasible. In addition, alternative consistent estimators of $\mathbf{H}^{\mathrm{MS}}$ can be constructed exploiting the specific structure of this example. One option is to employ a "plug-in" estimator of $\mathbf{H}^{\mathrm{MS}}$ based on nonparametric estimators of $f_{u|x_1,\mathbf{x}_2}$ and $f_{x_1|\mathbf{x}_2}$. An alternative, example-specific estimator is

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

where, for a differentiable kernel function K and a positive bandwidth h_n, $\dot{K}_n(u) = dK_n(u)/du$ and $K_n(u) = K(u/h_n)/h_n$. In words, $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ is constructed by "smoothing out" the indicator function entering $m^{\mathrm{MS}}(\mathbf{z}, \boldsymbol{\theta})$ and then twice differentiating the corresponding objective function (previously used by Horowitz (1992)).

4.2. *Panel Maximum Score*

Consider the panel data binary response model

$$Y_t = \mathbb{1}\big(\mathbf{X}_t'\boldsymbol{\beta}_0 + \alpha + u_t \geq 0\big), \quad t = 1, 2,$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^{d+1}$ is an unknown parameter of interest, α is an unobserved (time-invariant) individual-specific effect, and u_t is an unobserved error term. Analyzing this model, Manski (1987) gave conditions under which $\boldsymbol{\beta}_0$ is identified up to scale and demonstrated consistency of a conditional maximum score estimator.

Suppose $\boldsymbol{\beta}_0$ is identified up to scale and that its first element is positive, in which case we can normalize that element to unity and employ the parameterization $\boldsymbol{\beta}_0 = (1, \boldsymbol{\theta}_0')'$, where $\boldsymbol{\theta}_0 \in \mathbb{R}^d$ is unknown. To describe a version of the estimator of Manski (1987), partition $\mathbf{X}_t$ conformably with $\boldsymbol{\beta}_0$ as $\mathbf{X}_t = (X_{1t}, \mathbf{X}_{2t}')'$ and define $\mathbf{z} = (y, x_1, \mathbf{x}_2')'$, where $y = Y_2 - Y_1$,

$x_1 = X_{12} - X_{11}$, and $\mathbf{x}_2 = (\mathbf{X}_{22} - \mathbf{X}_{21})$. Assuming $\mathbf{z}_1, \dots, \mathbf{z}_n$ is a random sample of $\mathbf{z}$, a panel maximum score estimator of $\boldsymbol{\theta}_0$ is any $\hat{\boldsymbol{\theta}}_n^{\text{PMS}}$ approximately maximizing $\hat{M}_n$ for

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

As one would expect, the properties of $\hat{\boldsymbol{\theta}}_n^{\text{PMS}}$ are qualitatively similar to those of $\hat{\boldsymbol{\theta}}_n^{\text{MS}}$. To be specific, under regularity conditions (stated in Section A.3 of the Supplemental Material), the panel maximum score estimator is covered by the results of Section 3 and an example-specific alternative to the generic numerical derivative estimator is available, namely,

$$\tilde{\mathbf{H}}_n^{\text{PMS}} = -n^{-1} \sum_{i=1}^{n} y_i \dot{K}_n\big(x_{1i} + \mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n^{\text{PMS}}\big)\mathbf{x}_{2i}\mathbf{x}_{2i}',$$

where $\dot{K}_n$ is as in the maximum score example.

4.3. *Conditional Maximum Score*

Consider the dynamic panel data binary response model

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

where $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ and $\gamma_0 \in \mathbb{R}$ are unknown parameters of interest, α 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 γ_0 are identified up to a common scale factor. Assuming these conditions hold and that 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 $\mathbf{z}_1, \dots, \mathbf{z}_n$ is a random sample of $\mathbf{z}$, a conditional maximum score estimator of $\boldsymbol{\theta}_0$ is any $\hat{\boldsymbol{\theta}}_n^{\text{CMS}}$ approximately maximizing $\hat{M}_n$ for

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

where $\kappa_n(\mathbf{w}) = \kappa(\mathbf{w}/b_n)/b_n^d$ for a kernel function κ and a bandwidth b_n.

Through its dependence on b_n, the function m_n^{CMS} depends on n in a nonnegligible 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 $\hat{\boldsymbol{\theta}}_n^{\text{CMS}}$ turn out to be qualitatively similar to those of $\hat{\boldsymbol{\theta}}_n^{\text{MS}}$. To be specific, under regularity conditions (stated in Section A.4 of the Supplemental Material), the conditional maximum score estimator is covered by the results of Section 3 and an example-specific alternative to the generic numerical derivative estimator is available, namely,

$$\tilde{\mathbf{H}}_n^{\text{CMS}} = -n^{-1} \sum_{i=1}^{n} y_i \dot{K}_n\big(x_{1i} + \mathbf{x}_{2i}'\hat{\boldsymbol{\theta}}_n^{\text{CMS}}\big)\mathbf{x}_{2i}\mathbf{x}_{2i}'\kappa_n(\mathbf{w}_i),$$

where $\dot{K}_n$ is as in the maximum score example.

4.4. *Empirical Risk Minimization*

Mohammadi and van de Geer (2005) considered two-category classification problems in machine learning. Specifically, given a binary outcome $y \in \{-1, 1\}$ and a vector of features $\mathbf{x} \in \mathcal{X}$, the goal is to estimate the $\boldsymbol{\theta}_0$ that minimizes the misclassification error (or risk) $\mathbb{P}[h_{\boldsymbol{\theta}}(\mathbf{x}) \neq y]$ with respect to $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$, where $\{h_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is a collection of classifiers. For simplicity, we consider the case where the feature is univariate with support $\mathcal{X} = [0, 1]$ and the classifiers are of the form

$$h_{\boldsymbol{\theta}}(x) = \sum_{\ell=1}^{d+1} (-1)^{\ell} \mathbb{1}(\theta_{\ell-1} \leq x < \theta_{\ell}), \qquad \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)',$$

where $\boldsymbol{\Theta} = \{(\theta_1, \theta_2, \dots, \theta_d)' \in [0, 1]^d : 0 = \theta_0 \leq \theta_1 \leq \dots \leq \theta_d \leq \theta_{d+1} = 1\}$.

Assuming $\mathbf{z}_1, \dots, \mathbf{z}_n$ is a random sample of $\mathbf{z}$, an empirical risk minimizer is any $\hat{\boldsymbol{\theta}}_n^{\mathrm{ERM}}$ approximately maximizing $\hat{M}_n$ for $m_n(\mathbf{z}, \boldsymbol{\theta}) = m^{\mathrm{ERM}}(\mathbf{z}, \boldsymbol{\theta}) = -\mathbb{1}(h_{\boldsymbol{\theta}}(x) \neq y)$. Under regularity conditions similar to those of Mohammadi and van de Geer (2005, Section 2.1), the empirical risk minimizer is covered by Theorem 1 and the consistency requirement on $\tilde{\mathbf{H}}_n$ can be met in various ways; for details, see Section A.5 of the Supplemental Material.

5. SIMULATIONS

We illustrate the numerical performance of our proposed bootstrap-based inference methods for the maximum score estimator. Given the setup in Section 4.1, we generate data from that model with $d = 1$, $\boldsymbol{\theta}_0 = 1$, $\mathbf{x} = (x_1, x_2)' \sim \mathcal{N}((0, 1)', \mathbf{I}_2)$ with $\mathbf{I}_2$ the (2×2) identity matrix, and u generated by three distinct distributions. Specifically, DGP 1 sets $u \sim \mathsf{Logistic}(0, 1)/\sqrt{2\pi^2/3}$, DGP 2 sets $u \sim \mathcal{T}_3/\sqrt{3}$, where $\mathcal{T}_3$ denotes a Student's t-distribution with 3 degrees of freedom, and DGP 3 sets $u \sim (1 + 2(x_1 + x_2)^2 + (x_1 + x_2)^4)\mathsf{Logistic}(0, 1)/\sqrt{\pi^2/48}$.

The Monte Carlo experiment employs a sample size $n = 1000$ with $B = 2000$ bootstrap replications and $S = 2000$ simulations. For each of the three DGPs, we implement the standard nonparametric bootstrap, the m-out-of-n bootstrap using $m \in \{\lceil n^{1/2} \rceil, \lceil n^{2/3} \rceil, \lceil n^{4/5} \rceil\}$, and our proposed method using the two estimators $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ and $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ of $\mathbf{H}_0$. 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 (i) infeasible (simulation-based) MSE-optimal choices of tuning parameters (bandwidth/derivative step), denoted by $h_{\mathtt{MSE}}$ and $\epsilon_{\mathtt{MSE}}$, and (ii) infeasible and feasible AMSE-optimal choices of the tuning parameters, denoted by $h_{\mathtt{AMSE}}$, $\hat{h}_{\mathtt{AMSE}}$, $\epsilon_{\mathtt{AMSE}}$, and $\hat{\epsilon}_{\mathtt{AMSE}}$; for details, see Section A.2 of the Supplemental Material.

Table I 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 good finite sample performance in this Monte Carlo experiment. Results employing the example-specific plug-in estimator $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ are presented under the label "Plug-in" while results employing the generic numerical derivative estimator $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ are reported under the label "Num Deriv." Empirical coverage appears stable across different values of the tuning parameters for each method,

TABLE I

SIMULATIONS, MAXIMUM SCORE ESTIMATOR, 95% CONFIDENCE INTERVALS[a]

	DGP 1			DGP 2			DGP 3		
	h, ϵ	Coverage	Length	h, ϵ	Coverage	Length	h, ϵ	Coverage	Length
Standard									
		0.625	0.472		0.647	0.475		0.654	0.243
m-out-of-n									
$m = \lceil n^{1/2} \rceil$		0.997	1.698		0.998	1.753		1.000	1.890
$m = \lceil n^{2/3} \rceil$		0.978	1.185		0.983	1.221		0.989	0.724
$m = \lceil n^{4/5} \rceil$		0.899	0.820		0.897	0.837		0.930	0.447
Plug-in: $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$									
$h_{\mathtt{MSE}}$	0.620	0.954	0.511	0.580	0.957	0.523	0.150	0.962	0.277
$h_{\mathtt{AMSE}}$	1.108	0.972	0.590	0.480	0.951	0.518	0.123	0.942	0.263
$\hat{h}_{\mathtt{AMSE}}$	0.443	0.940	0.508	0.409	0.946	0.518	0.155	0.957	0.278
Num Deriv: $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$									
$\epsilon_{\mathtt{MSE}}$	1.400	0.936	0.483	1.360	0.938	0.485	0.290	0.939	0.249
$\epsilon_{\mathtt{AMSE}}$	0.537	0.880	0.414	0.573	0.894	0.426	0.224	0.902	0.227
$\hat{\epsilon}_{\mathtt{AMSE}}$	0.518	0.876	0.413	0.512	0.882	0.420	0.369	0.947	0.270

[a] (i) Panel *Standard* refers to standard nonparametric bootstrap, Panel *m-out-of-n* refers to m-out-of-n nonparametric bootstrap with subsample size m, Panel *Plug-in:* $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ refers to our proposed bootstrap-based method implemented using the example-specific plug-in drift estimator, and Panel *Num Deriv:* $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ refers to our proposed bootstrap-based method implemented using the generic numerical derivative drift estimator. (ii) Column "h, ϵ" 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 Material.

with better performance in the case of $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$. We conjecture that $n = 1000$ is too small for the numerical derivative estimator $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ to lead to as good inference performance as $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ (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 $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$ 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 $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$ implemented with the feasible bandwidth $\hat{h}_{\mathtt{AMSE}}$ perform quite well across the three DGPs considered.

6. CONCLUSION

We developed a valid resampling procedure for cube root asymptotics based on the nonparametric bootstrap. Whereas the standard nonparametric bootstrap is known to be invalid in the setting we study, we show that bootstrap validity can be restored by applying a carefully tailored reshapement of the objective function defining the estimator. Such reshapement is easy to implement both in general and in specific cases, as illustrated by the distinct examples we considered.

Seo and Otsu (2018) gave conditions under which results of the form (9) can be obtained also when the data exhibit weak dependence; see also Bagchi, Banerjee, and Stoev (2016), and references therein. It seems plausible that a version of our procedure, implemented with a resampling procedure suitable for dependent data, can be shown to be consistent under similar conditions, but it is beyond the scope of this paper to substantiate that conjecture.

7. PROOF OF THEOREM 1

The proof proceeds by first showing (9) and then using that result to establish (10). In both cases, we employ arguments similar to those used in the proof of the main theorem of Kim and Pollard (1990). The remainder of this section outlines the main steps in the proof; for technical details, see Lemmas A.1–A.10 in Section A.1 of the Supplemental Material.

Proof of (9). The estimator $\hat{\boldsymbol{\theta}}_n$ is assumed to satisfy

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

where

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

and

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

By the argmax continuous mapping theorem (e.g., van der Vaart and Wellner (1996), Theorem 3.2.2), it therefore suffices to show that $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ and that $\hat{G}_n + Q_n \rightsquigarrow \mathcal{G}_0 + \mathcal{Q}_0$ in the topology of uniform convergence on compacta. (The other conditions required by the argmax continuous mapping theorem are easily verified.)

To obtain the rate of convergence of $\hat{\boldsymbol{\theta}}_n$, we begin by using a standard argument to show that $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_{\mathbb{P}}(1)$ under Condition CRA(i) and then strengthen that conclusion to $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ by using Conditions CRA(ii)–(iii) and proceeding along the lines of van der Vaart and Wellner (1996, Theorem 3.2.5). In both cases, we employ the maximal inequality in Pollard (1989, Theorem 4.2); for details, see Lemmas A.1 and A.3 of the Supplemental Material.

Next, because Q_n is non-random, $\hat{G}_n + Q_n \rightsquigarrow \mathcal{G}_0 + \mathcal{Q}_0$ in the topology of uniform convergence on compacta if Q_n converges compactly to $\mathcal{Q}_0$ and if $\hat{G}_n \rightsquigarrow \mathcal{G}_0$ in the topology of uniform convergence on compacta. Compact convergence of Q_n follows from Condition CRA(ii); for details, see Lemma A.2 of the Supplemental Material. Also, to show that $\hat{G}_n \rightsquigarrow \mathcal{G}_0$ in the topology of uniform convergence on compacta, it suffices to show that $\hat{G}_n$ converges to $\mathcal{G}_0$ in the sense of weak convergence of finite-dimensional projections and that $\{\hat{G}_n(\mathbf{s}) : \|\mathbf{s}\| \leq K\}$ is stochastically equicontinuous for every $K > 0$.

Under Conditions CRA(ii)–(iv), weak convergence of finite-dimensional projections can be shown using the Cramér–Wold device and the fact that $\mathbb{E}[\hat{G}_n(\mathbf{s})\hat{G}_n(\mathbf{t})]$ converges to $\mathcal{C}_0(\mathbf{s}, \mathbf{t})$ for every $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$; for details, see Lemma A.4 of the Supplemental Material. Finally, under Conditions CRA(iii) and CRA(v) and employing the maximal inequality in Pollard (1989, Theorem 4.2), stochastic equicontinuity can be shown by proceeding as in the proof of Kim and Pollard (1990, Lemma 4.6); for details, see Lemma A.5 of the Supplemental Material.

Proof of (10). The proof of (10) is a natural bootstrap analog of the proof of (9). The estimator $\tilde{\boldsymbol{\theta}}_n^*$ is assumed to satisfy

$$\big\{\tilde{G}_n^*(\mathbf{s}) + \tilde{Q}_n(\mathbf{s})\big\}\big|_{\mathbf{s}=r_n(\tilde{\boldsymbol{\theta}}_n^*-\hat{\boldsymbol{\theta}}_n)} \geq \sup_{\mathbf{s}\in\mathbb{R}^d}\big\{\tilde{G}_n^*(\mathbf{s}) + \tilde{Q}_n(\mathbf{s})\big\} + o_{\mathbb{P}}(1),$$

where

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

and

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

By the argmax continuous mapping theorem, it therefore suffices to show that $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = O_{\mathbb{P}}(1)$ and that $\tilde{G}_n^* + \tilde{Q}_n \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0 + \mathcal{Q}_0$ in the topology of uniform convergence on compacta.

Using $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}_0$, to obtain the rate of convergence of $\tilde{\boldsymbol{\theta}}_n^*$ we first show that $\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n = o_{\mathbb{P}}(1)$ under Condition CRA(i) and then strengthen that conclusion to $r_n(\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n) = O_{\mathbb{P}}(1)$ by using $r_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_{\mathbb{P}}(1)$ and Condition CRA(iii). As in the derivation of the convergence rate of $\hat{\boldsymbol{\theta}}_n$, both steps employ the maximal inequality in Pollard (1989, Theorem 4.2); for details, see Lemmas A.6 and A.8 of the Supplemental Material.

Next, because $\mathcal{Q}_0$ is non-random, $\tilde{G}_n^* + \tilde{Q}_n \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0 + \mathcal{Q}_0$ in the topology of uniform convergence on compacta if $\tilde{Q}_n \to_{\mathbb{P}} \mathcal{Q}_0$ in the topology of uniform convergence on compacta and if $\hat{G}_n \rightsquigarrow \mathcal{G}_0$ in the topology of uniform convergence on compacta. By construction, $\tilde{Q}_n$ is such that if $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}_0$ and if $\hat{\boldsymbol{\theta}}_n \to_{\mathbb{P}} \boldsymbol{\theta}_0 \in \operatorname{int}(\boldsymbol{\Theta})$, then $\tilde{Q}_n \to_{\mathbb{P}} \mathcal{Q}_0$ in the topology of uniform convergence on compacta; for details, see Lemma A.7 of the Supplemental Material.

Also, to show that $\tilde{G}_n^* \rightsquigarrow_{\mathbb{P}} \mathcal{G}_0$ in the topology of uniform convergence on compacta, it suffices to show that $\tilde{G}_n^*$ converges to $\mathcal{G}_0$ in the sense of conditional weak convergence in probability of finite-dimensional projections and that $\{\tilde{G}_n^*(\mathbf{s}) : \|\mathbf{s}\| \leq K\}$ is stochastically equicontinuous for every $K > 0$. Conditional weak convergence in probability of finite-dimensional projections can be shown using the Cramér–Wold device and the fact that the maximal inequality in Pollard (1989, Theorem 4.2) can be used to show that $\mathbb{E}_n^*[\tilde{G}_n^*(\mathbf{s})\tilde{G}_n^*(\mathbf{t}))]$ converges in probability to $\mathcal{C}_0(\mathbf{s}, \mathbf{t})$ for every $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$, where $\mathbb{E}_n^*$ denotes an expectation computed under the bootstrap distribution conditional on the data; for details, see Lemma A.9 of the Supplemental Material. Finally, employing the maximal inequality in Pollard (1989, Theorem 4.2), stochastic equicontinuity can be shown by proceeding as in the proof of Kim and Pollard (1990, Lemma 4.6); for details, see Lemma A.10 of the Supplemental Material.

REFERENCES

ABREVAYA, J., AND J. HUANG (2005): "On the Bootstrap of the Maximum Score Estimator," *Econometrica*, 73 (4), 1175–1204. [2204,2206,2211]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78 (1), 119–157. [2204,2208]

BAGCHI, P., M. BANERJEE, AND S. A. STOEV (2016): "Inference for Monotone Functions Under Short-and Long-Range Dependence: Confidence Intervals and New Universal Limits," *Journal of the American Statistical Association*, 111 (516), 1634–1647. [2215]

BICKEL, P. J., AND B. LI (2006): "Regularization in Statistics," *Test*, 15 (2), 271–344. [2208]

BICKEL, P. J., F. GÖTZE, AND W. R. VAN ZWET (1997): "Resampling Fewer Than *n* Observations: Gains, Losses, and Remedies for Losses," *Statistica Sinica*, 7 (1), 1–31. [2204]

CATTANEO, M. D, M. JANSSON, AND K. NAGASAWA (2020): "Supplement to 'Bootstrap-Based Inference for Cube Root Asymptotics'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA17950. [2205]

CHERNOFF, H. (1964): "Estimation of the Mode," *Annals of the Institute of Statistical Mathematics*, 16 (1), 31–41. [2203,2208,2209]

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. [2204]

DÜMBGEN, L. (1993): "On Nondifferentiable Functions and the Bootstrap," *Probability Theory and Related Fields*, 95 (1), 125–140. [2204]

FANG, Z., AND A. SANTOS (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86 (1), 377–412. [2204,2208]

GRENANDER, U. (1956): "On the Theory of Mortality Measurement: Part II," *Scandinavian Actuarial Journal*, 39 (2), 125–153. [2203]

GROENEBOOM, P., AND K. HENDRICKX (2018): "Current Status Linear Regression," *Annals of Statistics*, 46 (4), 1415–1444. [2204]

GROENEBOOM, P., AND G. JONGBLOED (2018): "Some Developments in the Theory of Shape Constrained Inference," *Statistical Science*, 33 (4), 473–492. [2203]

HONG, H., AND J. LI (2020): "The Numerical Bootstrap," *Annals of Statistics*, 48 (1), 397–412. [2204]

HONORÉ, B. E., AND E. KYRIAZIDOU (2000): "Panel Data Discrete Choice Models With Lagged Dependent Variables," *Econometrica*, 68 (4), 839–874. [2203,2205,2208,2213]

HOROWITZ, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60 (3), 505–531. [2212]

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18 (1), 191–219. [2203,2205, 2209,2216,2217]

KOSOROK, M. R. (2008): "Bootstrapping the Grenander Estimator," in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*. Institute of Mathematical Statistics, 282–292. [2204, 2206]

LEE, S. M. S., AND M. C. PUN (2006): "On *m* out of *n* Bootstrapping for Nonstandard M-Estimation With Nuisance Parameters," *Journal of the American Statistical Association*, 101 (475), 1185–1197. [2204]

LEE, S. M. S., AND P. YANG (2020): "Bootstrap Confidence Regions Based on M-Estimators Under Nonstandard Conditions," *Annals of Statistics*, 48 (1), 274–299. [2204]

LÉGER, C., AND B. MACGIBBON (2006): "On the Bootstrap in Cube Root Asymptotics," *Canadian Journal of Statistics*, 34 (1), 29–44. [2204,2206]

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3 (3), 205–228. [2203,2205,2211]

——— (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27 (3), 313–333. [2205,2211]

——— (1987): "Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data," *Econometrica*, 55 (2), 357–362. [2205,2212]

MOHAMMADI, L., AND S. VAN DE GEER (2005): "Asymptotics in Empirical Risk Minimization," *Journal of Machine Learning Research*, 6, 2027–2047. [2203,2205,2214]

MUKHERJEE, D., M. BANERJEE, AND Y. RITOV (2019): "Non-Standard Asymptotics in High Dimensions: Manski's Maximum Score Estimator Revisited," arXiv:1903.10063. [2203]

PATRA, R. K., E. SEIJO, AND B. SEN (2018): "A Consistent Bootstrap Procedure for the Maximum Score Estimator," *Journal of Econometrics*, 205 (2), 488–507. [2204]

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. [2204]

POLLARD, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4 (4), 341–354. [2216,2217]

SEN, B., M. BANERJEE, AND M. WOODROOFE (2010): "Inconsistency of Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38 (4), 1953–1977. [2204,2206]

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. [2203-2205,2208,2209,2215]

SHI, C., W. LU, AND R. SONG (2018): "A Massive Data Framework for M-Estimators With Cubic-Rate," *Journal of the American Statistical Association*, 113 (524), 1698–1709. [2203]
VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer. [2207,2216]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 19 December, 2019; final version accepted 31 May, 2020; available online 26 June, 2020.

Econometrica Supplementary Material

SUPPLEMENT TO "BOOTSTRAP-BASED INFERENCE FOR CUBE ROOT ASYMPTOTICS"

(*Econometrica*, Vol. 88, No. 5, September 2020, 2203–2219)

MATIAS D. CATTANEO
Department of Operations Research and Financial Engineering, Princeton University

MICHAEL JANSSON
Department of Economics, University of California at Berkeley and CREATES

KENICHI NAGASAWA
Department of Economics, University of Warwick

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

APPENDIX

A.1. *Proofs of Main Results*

A.1.1. *Proof of Theorem 1*

AS EXPLAINED IN THE PAPER, Theorem 1 follows from ten technical lemmas. The remainder of this subsection presents those lemmas and their proofs.

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

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

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

PROOF OF LEMMA A.1: It suffices to show that every $\delta > 0$ admits a constant $c_\delta > 0$ such that

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

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

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

Matias D. Cattaneo: cattaneo@princeton.edu
Michael Jansson: mjansson@econ.berkeley.edu
Kenichi Nagasawa: Kenichi.Nagasawa@warwick.ac.uk

Cattaneo gratefully acknowledges financial support from the National Science Foundation through Grants SES-1459931 and SES-1947805, and Jansson gratefully acknowledges financial support from the National Science Foundation through Grants SES-1459967 and SES-1947662 and the research support of CREATES (funded by the Danish National Research Foundation under Grant DNRF78).

 https://doi.org/10.3982/ECTA17950

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

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

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

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

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

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

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

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

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

for any $K > 0$.

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

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

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

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

and

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

Q.E.D.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where the last inequality uses the Markov inequality.

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

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

which can be made arbitrarily small by making K large. *Q.E.D.*

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

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

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

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

the result follows from the Cramér–Wold device if

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

and if the following Lyapunov condition is satisfied:

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

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

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

we have

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

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

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

as was to be shown. *Q.E.D.*

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

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

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

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

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

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

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

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

and therefore

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

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

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

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

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

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

PROOF OF LEMMA A.6: It suffices to show that every $\delta > 0$ admits a constant $c_\delta^* > 0$ such that

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

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

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

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

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

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

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

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

Next, because

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

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

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

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

for any $K > 0$.

PROOF OF LEMMA A.7: Uniformly in $\mathbf{s}$ with $\|\mathbf{s}\| \leq K$, we have

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

where the last inequality uses $\tilde{\mathbf{H}}_n \to_{\mathbb{P}} \mathbf{H}$ and $\mathbb{P}(\hat{\boldsymbol{\theta}}_n + \mathbf{s}r_n^{-1} \notin \boldsymbol{\Theta}) \to 0$. *Q.E.D.*

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

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

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

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

$$\begin{aligned}
&\mathbb{P}\Big[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{M}_n^*(\boldsymbol{\theta}) - \tilde{M}_n^*(\tilde{\boldsymbol{\theta}}_n^*) \geq \delta r_n^{-2}\Big] + \mathbb{P}\big[\|\tilde{\mathbf{H}}_n - \mathbf{H}\| > \delta\big] + \mathbb{P}\big[\|\tilde{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n\| > \delta/4\big] \\
&\quad + \mathbb{P}\big[r_n\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| > 2^K\big] \\
&\quad + \sum_{j\geq K, 2^{j+1}\leq\delta r_n} \mathbb{P}\Big[\sup_{2^{j-1}<r_n\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n\|\leq 2^j, r_n\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\leq 2^K, \|\tilde{\mathbf{H}}_n-\mathbf{H}\|\leq\delta} \tilde{M}_n^*(\boldsymbol{\theta}) - \tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) \geq -\delta r_n^{-2}\Big].
\end{aligned}$$

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

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

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

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

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

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

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

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

we therefore have

$$\begin{aligned}
&\sum_{j\geq K, 2^{j+1}\leq\delta r_n} \mathbb{P}\Big[\sup_{2^{j-1}<r_n\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n\|\leq 2^j, r_n\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\leq 2^K, \|\tilde{\boldsymbol{H}}_n-\mathbf{H}\|\leq\delta} \tilde{M}_n^*(\boldsymbol{\theta}) - \tilde{M}_n^*(\hat{\boldsymbol{\theta}}_n) \geq -\delta r_n^{-2}\Big] \\
&\quad \leq \sum_{j\geq K, 2^{j+1}\leq\delta r_n} \mathbb{P}\Big[\sup_{2^{j-1}<r_n\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n\|\leq 2^j, r_n\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\leq 2^K} \big\{\hat{M}_n^*(\boldsymbol{\theta}) - \hat{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \hat{M}_n(\boldsymbol{\theta}) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n)\big\} \\
&\qquad \geq 2^{2j} c^* r_n^{-2}\Big] \\
&\quad \leq \sum_{j\geq K, 2^{j+1}\leq\delta r_n} \mathbb{P}\Big[\sup_{r_n\|\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_n\|\leq 2^j, r_n\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\|\leq 2^K} \big\|\hat{M}_n^*(\boldsymbol{\theta}) - \hat{M}_n^*(\hat{\boldsymbol{\theta}}_n) - \hat{M}_n(\boldsymbol{\theta}) + \hat{M}_n(\hat{\boldsymbol{\theta}}_n)\big\| \geq 2^{2j} c^* r_n^{-2}\Big]
\end{aligned}$$

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

where the last inequality uses the Markov inequality.

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

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

which can be made arbitrarily small by making K large. *Q.E.D.*

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

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

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

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

the result follows from the Cramér–Wold device if

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

and if the following Lyapunov condition is satisfied:

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

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

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

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),

$$r_n q_n \mathbb{E}_n^*\big[\big\{m_n\big(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}\big) - m_n\big(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n\big)\big\}\big\{m_n\big(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n + \mathbf{t} r_n^{-1}\big) - m_n\big(\mathbf{z}^*, \hat{\boldsymbol{\theta}}_n\big)\big\}\big] - \hat{\mathcal{C}}_n(\mathbf{s}, \mathbf{t})$$

$$= \frac{r_n q_n}{n} \sum_{i=1}^{n} \{ m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n + \mathbf{s} r_n^{-1}) - m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n) \} \{ m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n + \mathbf{t} r_n^{-1}) - m_n(\mathbf{z}_i, \hat{\boldsymbol{\theta}}_n) \} - \hat{\mathcal{C}}_n(\mathbf{s}, \mathbf{t})$$

$$= o_{\mathbb{P}}\left(r_n q_n \sqrt{\frac{r_n}{n q_n^3}} \right) = o_{\mathbb{P}}(1),$$

where

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

Using these facts and the representation

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

we have

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

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

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

Q.E.D.

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

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

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

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

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

where

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

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

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

and therefore,

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

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

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

A.1.2. *Proof of Lemma 1*

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

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

and

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

we obtain the decomposition

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

where

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

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

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

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

implying in particular that

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

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

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

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

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

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

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

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

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

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

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

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

A.1.3. *Proof of Lemma 2*

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

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

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

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

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

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

does not depend on ϵ_n.

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

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

where

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

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

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

Because

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

we have

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

Also, by condition CRA(iv),

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

Therefore,

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

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

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

A.1.4. *The Benchmark Case*

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

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

and

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

Condition CRA$_0$ (Cube Root Asymptotics, Benchmark Case) The following are satisfied:

(i) The class $\mathcal{M}_0$ 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, $\mathbf{H}_0 = -\partial^2 M_0(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is positive definite.

(iii) For some $\delta > 0$, the class $\{\mathcal{D}_0^{\delta'} : 0 < \delta' \leq \delta\}$ 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 $\delta_n > 0$ with $\delta_n = O(n^{-1/3})$, $n^{-1/3}\mathbb{E}[\bar{d}_0^{\delta_n}(\mathbf{z})^4] = o(1)$ 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 $\mathbf{s} \neq \mathbf{t}$,

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

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

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

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

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

A.2. *Example: Maximum Score*

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

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

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

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

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

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

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

is positive definite.

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

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

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

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

and

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

we have

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

and

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

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

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

Condition K The following are satisfied:

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

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

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

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

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

and

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

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

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

LEMMA MS: *Suppose Conditions MS and K hold.*

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

(ii) *If also $S_F \geq 3$ and $S_M \geq 4$, then $\tilde{H}_{n,kl}^{\mathtt{MS}}$ admits an approximation $\check{H}_{n,kl}^{\mathtt{MS}}$ satisfying*

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

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

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

A.2.1. *Proof of Corollary MS*

By Lemma A.11, it suffices to verify that Condition CRA$_0$ is satisfied.

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

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

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

Condition CRA$_0$(iv). Since $d_0^\delta(\mathbf{z})^4 = d_0^\delta(\mathbf{z})$, $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^4] = O(\delta_n)$, which implies the first condition. Also,

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

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

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

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

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

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

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

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

from which the desired result follows.

Condition CRA_0(v). The first part easily follows from $\bar{d}_0^{\delta}(\mathbf{z})\leq 1$, while the second part follows from the verification of Condition CRA_0(iv).

A.2.2. *Proof of Lemma MS*

A.2.2.1. *Part (i) [Consistency].* Defining

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

we obtain the decomposition

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

where

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

The proof will be completed by showing that $\check{\mathbf{H}}_n^{\mathtt{MS}}\to_{\mathbb{P}}\mathbf{H}^{\mathtt{MS}}$, $\mathbf{R}_n^{\mathtt{MS}}=o_{\mathbb{P}}(1)$, and $\mathbf{S}_n^{\mathtt{MS}}=o_{\mathbb{P}}(1)$. First, using the dominated convergence theorem and $\int_{\mathbb{R}}u\dot{K}(u)\,du=-1$, we have

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

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

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

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

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

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

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

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

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

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

Finally, defining

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

we have

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

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

A.2.2.2. *Part (ii) [Approximate MSE].* It was shown in the proof of part (i) that $R_{n,kl}^{\text{MS}} = o_{\mathbb{P}}(1/\sqrt{nh_n^3})$. Also, Taylor's theorem and Condition K(ii) can be used to show that for any

$C > 0$, we have, uniformly in $\|\boldsymbol{\delta}_n\| \leq C/\sqrt[3]{n}$,

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

implying in particular that

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

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

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

does not depend on h_n.

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

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

where

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

Also,

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

where

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

A.2.3. *Rule-of-Thumb Bandwidth Selection*

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

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

$F_0^{(2,2)}(\mathbf{x}_2) = 0,$

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

and

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

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

A.2.3.1. *Plug-in Estimator* $\tilde{\mathbf{H}}_n^{\mathrm{MS}}$. Given our reference model, natural estimators of the bias constants

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

are

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

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

Similarly, natural estimators of the variance constants

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

are given by

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

A.2.3.2. *Numerical Differentiation Estimator* $\tilde{\mathbf{H}}_n^{\mathrm{ND}}$. In our reference model, the bias constants are of the form

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

natural estimators of which are given by

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

Similarly, natural estimators of the variance constants

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

are given by

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

A.3. *Example: Panel Maximum Score*

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

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

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

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

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

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

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

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

is positive definite.

Letting

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

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

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

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

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

A.4. *Example: Conditional Maximum Score*

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

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

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

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

and

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

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

(i) For some strictly increasing F,

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

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

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

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

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

$$\mathbf{H}^{\text{CMS}}=\mathbb{E}\big[\eta^{\text{CMS}}(\mathbf{x}_2)\mathbf{x}_2\mathbf{x}_2'|\mathbf{w}\big]f_{\mathbf{w}}(\mathbf{w})|_{\mathbf{w}=\mathbf{0}}$$

is positive definite.

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

Let

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

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

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

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

A.4.1. *Proof of Corollary CMS*

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

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

Next, using the representations

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

we have

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

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

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

Condition CRA(ii). We have

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

and

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

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

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

Also, by a standard bias calculation for kernel estimators,

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

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

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

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

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

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

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

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

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

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

Finally, the representation

$$\begin{aligned}&\left\{m_n^{\text{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{s})-m_n^{\text{CMS}}(\mathbf{z},\boldsymbol{\theta})\right\}\left\{m_n^{\text{CMS}}(\mathbf{z},\boldsymbol{\theta}+\delta_n\mathbf{t})-m_n^{\text{CMS}}(\mathbf{z},\boldsymbol{\theta})\right\}\\&\quad=y^2\left[\mathbb{1}\left\{\delta_n\min\left(\mathbf{x}_2'\mathbf{s},\mathbf{x}_2'\mathbf{t}\right)\geq -x_1-\mathbf{x}_2'\boldsymbol{\theta}>0\right\}\right.\\&\qquad\left.+\mathbb{1}\left\{\delta_n\max\left(\mathbf{x}_2'\mathbf{s},\mathbf{x}_2'\mathbf{t}\right)<-x_1-\mathbf{x}_2'\boldsymbol{\theta}\leq 0\right\}\right]\kappa_n(\mathbf{w})^2\end{aligned}$$

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

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

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

A.5. *Example: Empirical Risk Minimization*

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

Condition ERM The following are satisfied:
- (i) $P(0) < 1/2$ and P admits a continuous derivative p in a neighborhood of each element of $\boldsymbol{\theta}_0$.
- (ii) F is absolutely continuous and its Lebesgue density f is continuously differentiable in a neighborhood of each element of $\boldsymbol{\theta}_0$.
- (iii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$.
- (iv) $\boldsymbol{\theta}_0 = (\theta_{0,1}, \theta_{0,2}, \dots, \theta_{0,d})'$ is the unique minimizer of $\mathbb{P}[h_{\boldsymbol{\theta}}(x) \neq y]$ and $p(\theta_{0,\ell}) \times f(\theta_{0,\ell}) \neq 0$ for $\ell \in \{1, \dots, d\}$.

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

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

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

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

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

$$\tilde{H}_{n,\ell\ell}^{\text{ERM}} = (-1)^{\ell+1} 2\hat{p}_n\big(\hat{\theta}_{n,\ell}^{\text{ERM}}\big)\hat{f}_n\big(\hat{\theta}_{n,\ell}^{\text{ERM}}\big), \quad \ell = 1, \dots, d,$$

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

A.5.1. *Proof of Corollary ERM*

By Lemma A.11, it suffices to verify that Condition CRA$_0$ is satisfied.

Condition CRA$_0$(i). Manageability of $\mathcal{M}_0$ follows from $\{\mathbb{1}(h_{\boldsymbol{\theta}}(x) \neq y) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ forming a VC subgraph class. Also, the envelope is bounded by 1. Finally, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_0^\delta} M_0(\boldsymbol{\theta}) < M_0(\boldsymbol{\theta}_0)$

for every $\delta > 0$ because $\boldsymbol{\Theta}$ is compact, M_0 is continuous, and $\boldsymbol{\theta}_0$ is the unique maximizer of $M_0(\boldsymbol{\theta})$.

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

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

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

Condition CRA_0(iii). This condition corresponds to the first part of (vii) in Theorem 7 of Mohammadi and van de Geer (2005).

Condition CRA_0(iv). Since $d_0^\delta(\mathbf{z})^4 = d_0^\delta(\mathbf{z})$, $\mathbb{E}[d_0^{\delta_n}(\mathbf{z})^4] = O(\delta_n)$, which implies the first condition. For the second part, Mohammadi and van de Geer (2005) showed that

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

Using the representations

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

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

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

As a consequence,

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

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

Condition CRA_0(v). The condition in display is identical to the second part of (vii) in Theorem 7 of Mohammadi and van de Geer (2005). The second assumption corresponds to (vi) in Theorem 7 of Mohammadi and van de Geer (2005).

REFERENCES

ABREVAYA, J., AND J. HUANG (2005): "On the Bootstrap of the Maximum Score Estimator," *Econometrica*, 73 (4), 1175–1204. [16]

HONORÉ, B. E., AND E. KYRIAZIDOU (2000): "Panel Data Discrete Choice Models With Lagged Dependent Variables," *Econometrica*, 68 (4), 839–874. [23]

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18 (1), 191–219. [5,9,16]

MOHAMMADI, L., AND S. VAN DE GEER (2005): "Asymptotics in Empirical Risk Minimization," *Journal of Machine Learning Research*, 6, 2027–2047. [24,25]

POLLARD, D. (1989): "Asymptotics via Empirical Processes," *Statistical Science*, 4 (4), 341–354. [1,4-6,8,11,12, 18]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 19 December, 2019; final version accepted 31 May, 2020; available online 26 June, 2020.