Bootstrapping Kernel-Based Semiparametric Estimators*

Matias D. Cattaneo

Department of Economics, University of Michigan

Michael Jansson

Department of Economics, UC Berkeley and *CREATES*

April 9, 2014

Abstract. This paper develops alternative asymptotic results for a large class of two-step semiparametric estimators. The first main result is an asymptotic distribution result for such estimators and differs from those obtained in earlier work on classes of semiparametric two-step estimators by accommodating a non-negligible bias. A noteworthy feature of the assumptions under which the result is obtained is that reliance on a commonly employed stochastic equicontinuity condition is avoided. The second main result shows that the bootstrap provides an automatic method of correcting for the bias even when it is non-negligible.

1. Introduction

This paper is concerned with basing inference about a finite-dimensional parameter on an estimator which is semiparametric in the sense that it employs a nonparametric

*We thank Stephane Bonhomme, Adam Rosen, Andres Santos, Azeem Shaikh, and seminar participants at University of Chicago, London School of Economics, Oxford University, Rice University, University College London, and the 2013 Latin American Meetings of the Econometric Society for comments. The first author gratefully acknowledges financial support from the National Science Foundation (SES 1122994). The second author gratefully acknowledges financial support from the National Science Foundation (SES 1124174) and the research support of *CREATES* (funded by the Danish National Research Foundation).

estimator of some nuisance function. The importance of such estimators is widely recognized, as is the difficulty of obtaining accurate distributional approximations for estimators of this kind. Regarding the latter, the consensus opinion (substantiated by Monte Carlo evidence) seems to be that the distributional properties of semiparametric estimators are much more sensitive to the properties of their (slowly converging) nonparametric ingredients than conventional asymptotic theory would suggest. In other words, the conventional approach to asymptotic analysis of semiparametric estimators, while delivering very tractable distributional approximations, effectively ignores certain features of these estimators which are important in samples of realistic size.

Important progress in the direction of avoiding this shortcoming has been made by Linton (1995), Nishiyama and Robinson (2000, 2001, 2005), and Ichimura and Linton (2005), among others. These papers develop (Nagar- and Edgeworth expansion-type) higher-order asymptotic theory under assumptions implying in particular that the estimators under consideration are $\sqrt{n}$-consistent (where n is the sample size). An alternative approach, and the one we take in this paper, was employed by Cattaneo, Crump, and Jansson (2013). That approach is conceptually similar to the "dimension asymptotics" approach taken in the seminal work of Mammen (1989), but we will refer to it as a "small bandwidth" approach for reasons that will become apparent below.[1,2] The "small bandwidth" approach is a first-order asymptotic approach whose goal is to produce more reliable distributional approximations by forcing the approximations to be more sensitive to the precise implementation of nonparametric ingredients than

[1]For an explanation of the connection between the approaches of Cattaneo et al. (2013) and Mammen (1989), see Enno Mammen's discussion of Cattaneo et al. (2013).

[2]The approach we take is also similar to the approach taken in a series of papers by Abadie and Imbens (2006, 2008, 2011), but our main conclusion regarding the bootstrap (and subsampling) is quite different from that of Abadie and Imbens (2008).

conventional first-order asymptotic approaches.

Because its objective is similar to that of asymptotic analysis based on Edgeworth expansions, an obvious question is whether the "small bandwidth" approach shares with the Edgeworth approach the feature that developing results for general classes of estimators is (or at least would appear to be) prohibitively complicated (e.g., Nishiyama and Robinson (2005, p. 927)). One of the two main goals of this paper is to demonstrate by example that this is not the case. To do so, we study a class of estimators essentially coinciding with the class investigated by Newey and McFadden (1994, Section 8) and develop "small bandwidth" asymptotic results for it. These results turn out to be in perfect qualitative agreement with those obtained by Cattaneo et al. (2013) for a particular member of the class of estimators under study. To be specific, it turns out that in general semiparametric estimators utilizing kernel estimators of unknown functions suffer from bias problems whose magnitude is non-negligible and bandwidth-dependent.

Being similar to that obtained by Cattaneo et al. (2013), this "small bandwidth" asymptotic finding is also in perfect analogy with the finding obtained under "dimension asymptotics" by Mammen (1989, Theorem 4). It therefore seems natural to ask whether positive results about the bootstrap analogous to those of Mammen (1989, Theorem 5) can be obtained also for the class of semiparametric estimators studied herein. Providing an affirmative answer to that question is the second main goal of this paper. Achieving this goal turns out to require relatively little effort, essentially because the high level conditions formulated in the process of achieving our first main goal have been designed partly with achievement of the second goal in mind.

From a practical perspective, our main methodological prescription is a simple and constructive one: in semiparametric models, inference procedures based on the bootstrap are much less sensitive to the precise implementation of nonparametric ingredients than their main rivals. Although this prescription is consistent with folklore, our theoretical justification for the prescription would appear to be new. In particular, unlike Nishiyama and Robinson (2005) our theory is based on first-order asymptotic results and partly for this reason we do not require studentization in order to show that the bootstrap provides "refinements" in the sense that bootstrap-based ("percentile") confidence intervals enjoy first-order asymptotic validity in cases where no such validity is enjoyed by their main rivals, including subsampling-based confidence intervals.

Previous work on bootstrap validity for general classes of semiparametric models includes Chen, Linton, and van Keilegom (2003) and Cheng and Huang (2010). For the models studied in this paper our results generalize theirs by accommodating non-parametric ingredients implemented using bandwidths that are "small" in the sense that they converge to zero at a faster-than-usual rate. Accommodating such bandwidths turns out to prohibit reliance on certain stochastic equicontinuity conditions employed by Chen et al. (2003), Cheng and Huang (2010), and most (if not all) previous developments of asymptotic distribution theory for semiparametric two-step estimators, including the results surveyed by Andrews (1994b), Newey and McFadden (1994), Chen (2007), and Ichimura and Todd (2007). As a consequence, avoiding reliance on such stochastic equicontinuity conditions turns out be necessary in order to achieve the goals of this paper. The approach taken in this paper is to replace a key stochastic equicontinuity condition by an "asymptotic separability" condition, which is similar to, but weaker than, its stochastic equicontinuity counterpart (in general). This condition exploits a special feature of kernel estimators, but may nevertheless be of independent interest.

The paper proceeds as follows. Section 2 provides a more detailed statement of and motivation for the questions addressed by this paper. Section 3 presents our "small bandwidth" asymptotic result about semiparametric two-step estimators, while Section 4 is concerned with verification of the high-level assumptions of that result. Section 5 presents bootstrap analogs of the results from Sections 3 and 4. Three appendices contain additional material.

2. Motivation

Suppose $\theta_0 \in \Theta \subseteq \mathbb{R}^k$ is a parameter of interest admitting an estimator $\hat{\theta}_n$ satisfying

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma), \tag{1}$$

where n is the sample size, $\rightsquigarrow$ denotes weak convergence (as $n \to \infty$), and Σ is some positive definite matrix. In this scenario it is common to base inference on a distributional approximation of the form $\sqrt{n}(\hat{\theta}_n - \theta_0)\dot{\sim}\mathcal{N}(0, \hat{\Sigma}_n)$, where $\hat{\Sigma}_n$ is some estimator of Σ. If $\hat{\Sigma}_n$ is consistent, then the distributional approximation is itself consistent in the sense that

$$\sup\nolimits_{t \in \mathbb{R}^k} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \leq t] - \mathbb{P}[\mathcal{N}(0, \hat{\Sigma}_n) \leq t] \right| \to_p 0, \tag{2}$$

a fact which in turn implies for instance that for any $\lambda \in \mathbb{R}^d$ the asymptotic coverage probability of the following confidence interval for $\lambda'\theta_0$ is 95%:

$$CI_n = \left[\lambda'\hat{\theta}_n - 1.96\sqrt{\lambda'\hat{\Sigma}_n\lambda/n}, \lambda'\hat{\theta}_n + 1.96\sqrt{\lambda'\hat{\Sigma}_n\lambda/n} \right].$$

An alternative distributional approximation is provided by the bootstrap. In standard notation, the bootstrap approximation to the cdf of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is given by $\mathbb{P}^*[\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \leq \cdot]$, where $\hat{\theta}_n^*$ denotes a bootstrap analogue of $\hat{\theta}_n$ and $\mathbb{P}^*$ denotes a probability computed under the bootstrap distribution conditional on the data. Assuming (1) holds, asymptotically valid inference procedures can be based on the bootstrap whenever the following bootstrap consistency condition is satisfied:

$$\sup_{t\in\mathbb{R}^k}\left|\mathbb{P}[\sqrt{n}(\hat{\theta}_n-\theta_0)\leq t]-\mathbb{P}^*[\sqrt{n}(\hat{\theta}_n^*-\hat{\theta}_n)\leq t]\right|\to_p 0. \tag{3}$$

For instance, for any $\lambda\in\mathbb{R}^d$ the asymptotic coverage probability of the following bootstrap-based confidence interval for $\lambda'\theta_0$ is 95%:

$$CI_n^*=\left[\lambda'\hat{\theta}_n-q_{n,0.975}^*,\lambda'\hat{\theta}_n-q_{n,0.025}^*\right],$$

where $q_{n,\alpha}^*=\inf\{q\in\mathbb{R}:\mathbb{P}^*[\lambda'(\hat{\theta}_n^*-\hat{\theta}_n)\leq q]\geq\alpha\}$.

Being "automatic" in the sense that it can be obtained without characterizing and/or (explicitly) estimating the matrix Σ, the bootstrap distributional approximation is particularly attractive when Σ is difficult to characterize and/or estimate, a phenomenon which occurs with some regularity for estimators $\hat{\theta}_n$ that are semiparametric in the sense that an estimator of an infinite-dimensional nuisance parameter is employed in the construction of $\hat{\theta}_n$. Chen et al. (2003) and Cheng and Huang (2010) made this point and gave conditions under which (1) and (3) are satisfied also in models with infinite-dimensional nuisance parameters.

In addition to complicating the characterization of Σ, the presence of estimators of infinite-dimensional nuisance parameters often implies that a delicate choice of tuning parameters (e.g., bandwidths in the case of kernel estimators) is required in order to achieve (1) in the first place. Moreover, in such semiparametric settings the finite sample distribution of $\sqrt{n}(\hat{\theta}_n-\theta_0)$ has often been found in Monte Carlo experiments to be rather sensitive to the choice of these tuning parameters, suggesting in particular that distributional approximations not depending on tuning parameters can be quite unreliable unless sample sizes are very large.

Acknowledging this, Cattaneo et al. (2013) investigated the consequences of relaxing the bandwidth conditions needed to achieve (1). Under weaker-than-usual bandwidth conditions and studying a particular kernel-based semiparametric estimator, Cattaneo et al. (2013) obtained a distributional result of the form

$$\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathcal{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma), \tag{4}$$

where Σ is the same as in (1) while $\mathcal{B}_n$ is some (possibly) non-negligible bias whose value depends in part on the bandwidth. Simulation evidence reported by Cattaneo et al. (2013) was found there to be consistent with the main prediction obtained by replacing (1) with the more general result (4), namely that kernel-based semiparametric estimators suffer from bias problems whose magnitude is non-negligible and bandwidth-dependent.

Replacing (1) with (4) can have severe consequences. For instance, the consistency property (2) fails when $\mathcal{B}_n \neq o(n^{-1/2})$ in (4), implying in turn that inference procedures based on the distributional approximation $\sqrt{n}(\hat{\theta}_n - \theta_0) \dot{\sim} \mathcal{N}(0, \hat{\Sigma}_n)$ are invalid in general.[3] On the other hand, using the relation

$$\begin{aligned} & \sup_{t \in \mathbb{R}^k} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathcal{B}_n) \leq t] - \mathbb{P}^*[\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n - \mathcal{B}_n) \leq t] \right| \\ = \; & \sup_{t \in \mathbb{R}^k} \left| \mathbb{P}[\sqrt{n}(\hat{\theta}_n - \theta_0) \leq t] - \mathbb{P}^*[\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \leq t] \right| \end{aligned} \tag{5}$$

it can be shown that (3) and (4) are sufficient to guarantee asymptotic validity of inference procedures (such as the 95% confidence interval CI^*_n) based on the bootstrap.

[3] For instance, the asymptotic coverage probability of the interval CI_n is less than 95% when $\lambda' \mathcal{B}_n \neq o(n^{-1/2})$ (and equal to zero when $\lambda' \mathcal{B}_n \neq O(n^{-1/2})$).

For bootstrap-based inference procedures the consequences of replacing (1) with (4) would therefore be benign if validity of (3) could be established also under (4) .

The objective of this paper is twofold. First, we want to explore the generality of the finding that replacing (1) with (4) is necessary when characterizing the large-sample properties of semiparametric estimators under weaker-than-usual conditions on tuning parameters. To do so, we study an important class of kernel-based semiparametric two-step estimators and find that members of this class of estimators generally have the property that a result of the form (4) can be obtained under significantly weaker bandwidth conditions than those needed to achieve (1) .

Our development is constructive in the sense that it produces an explicit and interpretable formula for the bias $\mathcal{B}_n$. It turns out that $\mathcal{B}_n$ in (4) arises due to features of $\hat{\theta}_n$ that can be replicated by the bootstrap. As a consequence, it seems plausible that the bootstrap consistency property (3) could be valid also when $\mathcal{B}_n \neq o(n^{-1/2})$ in (4) . Formalizing and verifying the latter conjecture is the second objective of this paper. In combination, our findings suggest that even though semiparametric estimators are likely to suffer from nonnegligible bias problems in samples of moderate size, these biases can be corrected for in a fully automatic way by basing inference on the bootstrap. It seems to us that this is a "robustness" property of the bootstrap that is of both theoretical and practical importance.

Remarks. (i) Although our main emphasis is on obtaining constructive results, one negative conclusion emerging as a by-product of our development seems to be of sufficient theoretical and practical interest to be worth mentioning. As it turns out, the "robustness" of bootstrap-based inference with respect to tuning parameter choice is not shared by inference procedures based on subsampling, a possibly surprising finding in light of the fact that subsampling is often regarded as a "regularized" version of the bootstrap (e.g., Bickel and Li (2006)). Indeed, while it is well understood (e.g., Politis and Romano (1994)) that the subsampling approximation to the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is consistent under (1) whenever $\hat{\theta}_n$ is of the form $\hat{\theta}_n = T_n(z_1, \ldots, z_n)$ with $z_i \sim i.i.d.$, a consequence of of replacing (1) with (4) is that one of the "minimal" assumptions for asymptotic validity of subsampling is violated (in general) and in fact it is not hard to show that inference procedures based on subsampling will be invalid (in general) as well. (For instance, subsampling-based inference procedures will be invalid if, for some $r < 1/2$, $n^r \mathcal{B}_n$ converges to a non-zero limit.) In particular, the examples analyzed below all have the feature that there are (bandwidth) conditions under which inference procedures based on the standard bootstrap are valid even though subsampling-based inference procedures are not.

(ii) Ibragimov and Müller (2010) have proposed an inference procedure which shares with subsampling-based inference procedures the feature that its asymptotic validity follows from (1) whenever $\hat{\theta}_n$ is of the form $\hat{\theta}_n = T_n(z_1, \ldots, z_n)$ with $z_i \sim i.i.d.$. Like subsampling, that procedure ceases to be valid (in general) when (1) is replaced by (4).

3. Asymptotics without Stochastic Equicontinuity

Suppose the estimand θ_0 is the solution to an equation of the form $G(\theta, \gamma_0) = 0$, where $G(\theta, \gamma) = \mathbb{E}g(z, \theta, \gamma(\cdot, \theta))$, $g(\cdot)$ is a known function, z is a random vector, and

$$\gamma_0(z, \theta) = \mathbb{E}[w(z, \theta)|x(z, \theta)] f_0[x(z, \theta), \theta],$$

with $w(\cdot)$ and $x(\cdot)$ being known functions and $f_0(\cdot, \theta)$ denoting the (unknown) density of $x(z, \theta)$, the latter being assumed to be continuously distributed.

Letting $z_1, \ldots, z_n$ denote $i.i.d.$ copies of z, a natural estimator $\hat{\theta}_n$ of θ_0 is given by an approximate minimizer (with respect to $\theta \in \Theta$, Θ being the parameter space) of

$$\hat{G}_n(\theta, \hat{\gamma}_n)' W \hat{G}_n(\theta, \hat{\gamma}_n), \qquad \hat{G}_n(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} g[z_i, \theta, \gamma(\cdot, \theta)],$$

where W is some symmetric, positive semi-definite matrix and $\hat{\gamma}_n$ is a kernel-based estimator of γ_0 given by

$$\hat{\gamma}_n(z, \theta) = \frac{1}{n} \sum_{j=1}^{n} w(z_j, \theta) \mathcal{K}_n[x(z, \theta) - x(z_j, \theta)], \qquad \mathcal{K}_n(x) = \frac{1}{h_n^d} K\left(\frac{x}{h_n}\right),$$

where $K(\cdot)$ is a kernel and d is the dimension of $x(z, \theta)$.[4]

The formulation just given is essentially the same as in Newey and McFadden (1994, Section 8), except that we follow Newey (1994) and Chen et al. (2003), respectively, by allowing the dimension of g to exceed that of θ_0 and by accommodating profiling (i.e., allowing $\gamma_0(z, \theta)$ and $\hat{\gamma}_n(z, \theta)$ to depend on θ) as well as models where $g(z, \cdot, \cdot)$ is non-smooth.

3.1. Conventional Asymptotics. Theorem 2 below provides a template for establishing (4). The approach summarized in that theorem is related to Newey and McFadden's (1994) approach to establishing (1). To facilitate a comparison between the approaches, we begin by presenting a version of that approach.

Lemma 1. *Suppose that:*

(AL) [approximate linearity] for some matrix $\mathcal{J}$ of rank k,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{J} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_0(z_i, \hat{\gamma}_n) + o_p(1),$$

[4] In our motivating examples, an (approximate) minimizer is one that solves $\hat{G}_n(\theta, \hat{\gamma}_n) = 0$. More generally, an (approximate) minimizer is one that satisfies Condition (i) of Lemma 3.

where $g_0(z,\gamma) = g[z,\theta_0,\gamma(\cdot,\theta_0)]$;

(SE) [stochastic equicontinuity] for some function $\bar{g}_0$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_0(z_i,\hat{\gamma}_n) - \bar{g}_0(z_i,\hat{\gamma}_n) - g_0(z_i,\gamma_0) + \bar{g}_0(z_i,\gamma_0)] \to_p 0,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_0(z_i,\hat{\gamma}_n) - \bar{G}_0(\hat{\gamma}_n) - \bar{g}_0(z_i,\gamma_0) + \bar{G}_0(\gamma_0)] \to_p 0,$$

where $\bar{G}_0(\gamma) = \mathbb{E}\bar{g}_0(z,\gamma)$;

(AN$_0$) [asymptotic zero-mean normality] for some positive definite Ω,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_0(z_i,\gamma_0) + \bar{G}_0(\hat{\gamma}_n) - \bar{G}_0(\gamma_0)] \rightsquigarrow \mathcal{N}(0,\Omega).$$

Then (1) *holds with* $\Sigma = \mathcal{J}\Omega\mathcal{J}'$.

Condition (AL) is standard and typically holds (with $\mathcal{J} = -(\dot{G}_0' W \dot{G}_0)^{-1}\dot{G}_0' W$) provided the error in the following linear approximation to $\hat{G}_n(\cdot,\hat{\gamma}_n)$ is small:

$$\hat{G}_n(\theta,\hat{\gamma}_n) \approx \hat{G}_n(\theta_0,\hat{\gamma}_n) + \dot{G}_0(\theta-\theta_0), \qquad \dot{G}_0 = \frac{\partial}{\partial\theta'}G(\theta,\gamma_0)\bigg|_{\theta=\theta_0}.$$

A sufficient condition for this occur will be given in Section 4.1.

An implication of Condition (AL) is that the large sample properties of $\hat{\theta}_n$ are governed by $n^{-1/2}\sum_{i=1}^{n} g_0(z_i,\hat{\gamma}_n)$. Characterizing the distributional properties of such objects is potentially complicated because of the possible nonlinearity of $g_0(z_i,\cdot)$ and, in particular, the dependence/overlap between the arguments of $g_0(z_i,\hat{\gamma}_n)$ (i.e., between z_i and $\hat{\gamma}_n$). A common way to account for nonlinearity and the overlap is to impose Condition (SE) with $\bar{g}_0(z_i,\cdot)$ being a linear approximation to $g_0(z_i,\cdot)$. Irrespective of the functional form of $\bar{g}_0$, Condition (SE) implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_0(z_i,\hat{\gamma}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_0(z_i,\gamma_0) + \bar{G}_0(\hat{\gamma}_n) - \bar{G}_0(\gamma_0)] + o_p(1). \tag{6}$$

Because the summands in this expansion depend on either z_i or $\hat{\gamma}_n$ (but not both), Condition (SE) effectively achieves (asymptotic) "separability" between z_i and $\hat{\gamma}_n$.

Verifying Condition (AN$_0$) on the part of the leading term in the expansion (6) tends to be straightforward, as typically the $g_0(z_i, \gamma_0)$ are mean zero random variables and $\bar{G}_0(\hat{\gamma}_n) - \bar{G}_0(\gamma_0)$ is a smooth functional of $\hat{\gamma}_n$. (For instance, $\bar{G}_0$ is linear whenever $\bar{g}_0(z_i, \cdot)$ is.) To be specific, assuming the smoothing bias of $\hat{\gamma}_n$ is small enough it is usually not hard to show that Condition (AN$_0$) holds (with Ω computable using the pathwise derivative formula of Newey (1994)).

Remarks. (i) From the perspective of this paper the most problematic assumption in Lemma 1 is Condition (SE). One exceptional case where that condition is mild is the case where the model is "adaptive" in the sense that the first step (i.e., estimation of γ_0) has no effect on the asymptotic distribution of $\hat{\theta}_n$. This occurs when

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g_0(z_i, \hat{\gamma}_n) - g_0(z_i, \gamma_0)] \to_p 0, \tag{7}$$

in which case Condition (SE) is satisfied with $\bar{g}_0 = 0$ (and Condition (AN$_0$) is satisfied with $\Omega = \mathbb{V}[g_0(z_i, \gamma_0)]$ whenever the latter exists and is positive definite). In the sequel we tacitly assume that we are dealing with a "non-adaptive" situation where the first step does have an effect on the asymptotic distribution of $\hat{\theta}_n$. This assumption is less restrictive than might at first appear to be the case, as it turns out that achieving (7) under the "small bandwidth" asymptotics employed in this paper requires somewhat stronger conditions than in cases where the nonparametric ingredient $\hat{\gamma}_n$ is $n^{1/4}$-consistent. For details, see Section 7.1.

(ii) When $\bar{g}_0(z_i, \cdot)$ is linear, the first part of Condition (SE) typically holds provided $\hat{\gamma}_n - \gamma_0 = o_p(n^{-1/4})$. More generally, when $\bar{g}_0(z_i, \cdot)$ is a polynomial approximation of order R, then the first part of Condition (SE) typically holds provided $\hat{\gamma}_n - \gamma_0 = o_p(n^{-1/2(R+1)})$. In other words, the first part of Condition (SE) can typically be satisfied by judicious choice of $\bar{g}_0$. (Indeed, the first part of Condition (SE) can be rendered redundant by setting $\bar{g}_0 = g_0$.) The most interesting part of Condition (SE), and the part motivating the label "(SE)", is therefore the second part, which is a stochastic equicontinuity condition.

(iii) In important special cases, the second part of Condition (SE) reduces to conditions already in the literature, being equivalent to Assumption 5.2 of Newey (1994) when $\bar{g}_0(z_i, \cdot)$ is linear and reducing to (2.8) of Andrews (1994a) and (3.34) of Andrews (1994b) when $\bar{g}_0 = g_0$.

3.2. Alternative Asymptotics. As can the approach outlined in Lemma 1, the approach taken in this paper can be summarized by means of three high-level conditions. To state these, let

$$\hat{\gamma}_{n,i}(z, \theta) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} w(z_j, \theta) \mathcal{K}_n[x(z, \theta) - x(z_j, \theta)] \qquad (i = 1, \ldots, n)$$

denote the "leave-one-out" versions of $\hat{\gamma}_n(z, \theta)$ and define $\gamma_n(\cdot, \theta) = \mathbb{E}\hat{\gamma}_n(\cdot, \theta)$.

Theorem 2. *Suppose that:*

(AL) for some matrix $\mathcal{J}$ of rank k,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \mathcal{J} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_n(z_i, \hat{\gamma}_{n,i}) + o_p(1),$$

where $g_n(z, \gamma) = g[z, \theta_0, n^{-1} w(z, \theta_0) \mathcal{K}_n[x(\cdot, \theta_0) - x(z, \theta_0)] + (1 - n^{-1})\gamma(\cdot, \theta_0)]$;

(AS) [asymptotic separability] for some function $\bar{g}_n$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\hat{\gamma}_{n,i})-\bar{g}_n(z_i,\hat{\gamma}_{n,i})-g_n(z_i,\gamma_n)+\bar{g}_n(z_i,\gamma_n)]\to_p 0,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_n(z_i,\hat{\gamma}_{n,i})-\bar{G}_n(\hat{\gamma}_{n,i})-\bar{g}_n(z_i,\gamma_n)+\bar{G}_n(\gamma_n)]\to_p 0,$$

where $\bar{G}_n(\gamma)=\mathbb{E}\bar{g}_n(z,\gamma)$;

(AN) [asymptotic normality] for some positive definite Ω *and some* β_n,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\gamma_n)+\bar{G}_n(\hat{\gamma}_{n,i})-\bar{G}_n(\gamma_n)-\beta_n]\rightsquigarrow\mathcal{N}(0,\Omega).$$

Then (4) *holds with* $\mathcal{B}_n=\mathcal{J}\beta_n$ *and* $\Sigma=\mathcal{J}\Omega\mathcal{J}'$.

By construction, the functional form of g_n is such that $g_n(z_i,\hat{\gamma}_{n,i})=g_0(z_i,\hat{\gamma}_n)$ for every $i=1,\ldots,n$. As a consequence, Condition (AL) of the theorem is simply an unorthodox restatement of Condition (AL) of Lemma 1. It is stated in terms of g_n and $\hat{\gamma}_{n,i}$ in anticipation of the other conditions of the theorem.

As discussed in more detail in Section 4.2, Condition (SE) typically fails when the bandwidth h_n is "small" in the sense that it converges to zero at a faster-than-usual rate. On the one hand, rapid convergence of h_n to zero can result in failure of $n^{1/4}$-consistency of $\hat{\gamma}_n$, which in turn can lead to a failure of the first part of Condition (SE) when $\bar{g}_0(z_i,\cdot)$ is chosen to be linear. In isolation, this problem is conceptually straightforward to address. Indeed, as alluded to in remark (ii) at the end of the previous subsection, the first part of Condition (SE) can usually be preserved also when smaller-than-usual bandwidths are employed simply by working with a polynomial (e.g., quadratic) $\bar{g}_0(z_i,\cdot)$. On the other hand, it turns out that regardless of the functional form of $\bar{g}_0(z_i,\cdot)$ the stochastic equicontinuity (i.e., the second) part of Condition (SE) tends to fail when the bandwidth conditions ensuring $n^{1/4}$-consistency of $\hat{\gamma}_n$ are relaxed. Addressing this problem turns out to be necessary in order to achieve the goals of this paper.

We are unaware of previous work pointing out the need to (let alone providing a solution to the question of how to) avoid reliance on stochastic equicontinuity when accommodating slowly converging nonparametric components. The approach we take is to replace Condition (SE) with Condition (AS). On the surface, Condition (AS) is simply a "leave-one-out" version of Condition (SE), but perhaps remarkably it turns out that Condition (AS) is satisfied (with $\bar{g}_n(z_i, \cdot)$ being a quadratic approximation to $g_n(z_i, \cdot)$) in many cases even when Condition (SE) is not.

Condition (AS) gives rise to the following counterpart of (6) :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_n(z_i, \hat{\gamma}_{n,i}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g_n(z_i, \gamma_n) + \bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n)] + o_p(1). \tag{8}$$

In this expansion, the terms $g_n(z_i, \gamma_n)$ and $\bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n)$ each depend on one (but not both) of z_i and $\hat{\gamma}_{n,i}$. In other words, Condition (AS) achieves (asymptotic) "separability" between z_i and the nonparametric ingredient $\hat{\gamma}_{n,i}$ in $g_n(z_i, \hat{\gamma}_{n,i})$. As such, Condition (AS) serves a purpose very similar to that of Condition (SE) and the label "(AS)" has been chosen to highlight this connection between the two conditions.

The big advantage of Condition (AS) is that in leading examples it can be verified under assumptions that are considerably weaker than those required for Condition (SE). The price to be paid for this extra generality is relatively small. In addition to the notational nuisance of having to employ additional subscripts in many places, a complication that must be addressed is that the terms $g_n(z_i, \gamma_n)$ and $\bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n)$ in (8) both have a nonnegligible mean in general. Accordingly, Condition (AN) is the simplest counterpart of Condition (AN_0) that one can hope for (in general). Thankfully, it turns out that Condition (AN) is verifiable under assumptions similar to those required for Condition (AN_0) (with Ω once again computable using the pathwise derivative formula of Newey (1994) and β_n given by a formula presented below).

4. Verifying the Conditions of Theorem 2

The goal in this section is to demonstrate the plausibility of the conditions of Theorem 2. To do so, we outline general approaches to verifying the conditions and apply these approaches to the following three prominent examples.[5]

Example 1: ***Average Density.*** Suppose $x_1, \ldots, x_n$ are *i.i.d.* copies of a continuously distributed random vector $x \in \mathbb{R}^d$ with density f_0. Then a kernel-based estimator of $\theta_0 = \int_{\mathbb{R}^d} f_0(x)^2 dx$, the average density, is given by

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_n(x_i), \qquad \hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathcal{K}_n(x - x_j).$$

When verifying the conditions of Theorem 2 for this example, we set $z = x$, $x(z, \theta) = z$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = f_0(\cdot)$, and let $\hat{\theta}_n$ be defined by $\hat{G}_n(\hat{\theta}_n, \hat{f}_n) = 0$, where $g(x, \theta, f) = f(x) - \theta$ is a linear functional of f.

Being a second-order V-statistic this estimator is very tractable. Owing partly to this tractability the estimator has been widely studied (in the statistics literature at least). We include it here mainly because it provides a dramatic demonstration of the fragility of the second part of Condition (SE) with respect to bandwidth choice. ■

Example 2: ***Weighted Average Derivative.*** Suppose $z_1, \ldots, z_n$ are *i.i.d.* copies of $z = (y, x')'$, where $y \in \mathbb{R}$ is a scalar dependent variable and the vector $x \in \mathbb{R}^d$ is a continuous explanatory variable with density f_0. A weighted average derivative of the regression function $r(x) = \mathbb{E}(y|x)$ is defined as $\theta_0 = \mathbb{E}[\omega(x)\partial r(x)/\partial x]$, where

[5] To avoid distractions, we omit statements of regularity conditions when presenting the examples and the associated results. Additional details for the examples are provided in Section 9.

ω is a known scalar weight function. As an estimator of θ_0, Cattaneo et al. (2013) considered

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} y_i s(x_i, \hat{f}_n), \qquad s(x, f) = -\frac{\partial}{\partial x}\omega(x) - \omega(x)\frac{\partial f(x)/\partial x}{f(x)},$$

where $\hat{f}_n$ is defined as in Example 1.

In this example, $z = (y, x')'$, $x(z, \theta) = x$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = f_0(\cdot)$, and $\hat{G}_n(\hat{\theta}_n, \hat{f}_n) = 0$, where $g(z, \theta, f) = ys(x, f) - \theta$.

This estimator is a representative member of the class of two-step semiparametric estimators insofar as it involves a nonlinear, but smooth, functional of its nonparametric ingredient $\hat{f}_n$. As will become apparent, this nonlinearity needs to be taken into account when choosing $\bar{g}_n$ in the process of verifying Condition (AS). More importantly, perhaps, it turns out that nonlinearity manifests itself in the functional form of β_n in Condition (AN) and therefore in the form of the bias $\mathcal{B}_n$ in (4). ■

Example 3: ***Hit Rate.*** Suppose $z_1, \ldots, z_n$ are *i.i.d.* copies of $z = (y, x')'$, where $y \in \mathbb{R}$ is a scalar and the vector $x \in \mathbb{R}^d$ is a continuous explanatory variable with density f_0. As an estimator of $\theta_0 = \mathbb{P}[y \geq f_0(x)]$, a particular example of a so-called 'hit rate', Chen et al. (2003) studied

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} 1[y_i \geq \hat{f}_n(x_i)],$$

where $1[\cdot]$ is the indicator function and $\hat{f}_n$ is as in the previous examples.

In this example, $z = (y, x')'$, $x(z, \theta) = x$, $w(z, \theta) = 1$, $\gamma_0(\cdot, \theta) = f_0(\cdot)$, and $\hat{G}_n(\hat{\theta}_n, \hat{f}_n) = 0$, where $g(z, \theta, f) = 1[y \geq f(z)] - \theta$.

While simple in many respects, being a discontinuous functional of $\hat{f}_n$ this estimator is an interesting one to investigate when attempting to understand the extent to which the conclusions of Example 2 are consequences of the smoothness (with respect to $\hat{f}_n$) of the estimator considered there. In particular, because the empirical process methods utilized by Chen et al. (2003) to handle the lack of smoothness of $\hat{\theta}_n$ with respect to $\hat{f}_n$ cannot be applied when verifying Condition (AS) it would appear to be of interest to investigate the extent to which Condition (AS) can be verified also in the absence of smoothness. ■

4.1. Approximate Linearity. Condition (AL) is simply a representation when $\hat{\theta}_n$ is defined as the solution to $\hat{G}_n(\theta, \hat{\gamma}_n) = 0$ for a function g with $g(z, \theta, \gamma) - \theta$ not depending on θ. Indeed, Condition (AL) holds with $\mathcal{J} = I_k$ and without any $o_p(1)$ term in this important special case, which covers Examples 1-3.

More generally, verifying Condition (AL) is usually straightforward when $\hat{G}_n(\theta_0, \hat{\gamma}_n)$ is assumed to be $O_p(n^{-1/2})$, as in Conditions (SE) and (AN$_0$) of Lemma 1, because then sufficient conditions for Condition (AL) can be formulated with the help of Pakes and Pollard (1989, Theorem 3.3). The situation is slightly more delicate in the scenarios of main interest to us because Conditions (AS) and (AN) of Theorem 2 imply only that $\hat{G}_n(\theta_0, \hat{\gamma}_n) = O_p(\|\beta_n\|)$, where it turns out that $\|\beta_n\| = o(n^{-1/3}) \neq O(n^{-1/2})$ in the cases of primary interest.

For completeness, we provide a set of sufficient conditions for Condition (AL) compatible with Conditions (AS) and (AN). To state these, let $\|\cdot\|$ denote the Euclidian norm, let $\|\cdot\|_\Gamma$ denote a norm on a function space to which $\hat{\gamma}_n - \gamma_0$ belongs (with probability approaching one), and for any $\delta > 0$, let $\Theta(\delta) = \{\theta : \|\theta - \theta_0\| \leq \delta\}$ and $\Gamma(\delta) = \{\gamma : \|\gamma - \gamma_0\|_\Gamma \leq \delta\}$. Also, suppose $\dot{G}(\gamma) = \partial G(\theta, \gamma)/\partial\theta'|_{\theta=\theta_0}$ exists for every γ in a neighborhood of γ_0 and let $\dot{G}_0 = \dot{G}(\gamma_0)$.

Lemma 3. *Suppose that* $\hat{\theta}_n - \theta_0 = o_p(1)$, $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, *and that:*
(i) $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \leq \inf_{\theta \in \Theta} \hat{G}_n(\theta, \hat{\gamma}_n)' W \hat{G}_n(\theta, \hat{\gamma}_n) + o_p(n^{-1})$;

(ii) $\dot{G}_0' W \dot{G}_0$ *has rank* k *and for every positive* $\delta_n = o(1)$,

$$\sup_{\theta \in \Theta(\delta_n), \gamma \in \Gamma(\delta_n)} \left[\frac{\|G(\theta,\gamma) - G(\theta_0,\gamma) - \dot{G}(\gamma)(\theta-\theta_0)\|}{\|\theta-\theta_0\|^2} + \frac{\|\dot{G}(\gamma) - \dot{G}_0\|}{\|\gamma-\gamma_0\|_\Gamma} \right] = O(1);$$

(iii) for $\alpha \in \{0, 1/3\}$ *and for every positive* $\delta_n = o(n^{-\alpha})$,

$$\sup_{\theta \in \Theta(\delta_n)} \|\hat{G}_n(\theta,\hat{\gamma}_n) - G(\theta,\hat{\gamma}_n) - \hat{G}_n(\theta_0,\hat{\gamma}_n) + G(\theta_0,\hat{\gamma}_n)\| = o_p(n^{-1/2-\alpha/2});$$

(iv) $\hat{G}_n(\theta_0,\hat{\gamma}_n) = o_p(n^{-1/3})$;

(v) θ_0 *is an interior point of* Θ.

Then Condition (AL) holds with $\mathcal{J} = -(\dot{G}_0' W \dot{G}_0)^{-1} \dot{G}_0' W$.

Lemma 3 is in the spirit of Pakes and Pollard (1989, Theorem 3.3). For more details on the assumptions of the lemma, including in particular some remarks on our (stochastic equicontinuity) condition (iii), see the discussion in Section 7.2.

Suffice it to say that overall the conditions of Lemma 3 seem sufficiently weak to support the view that in perfect analogy Newey and McFadden (1994, p. 2196), Condition (AL) is “not conceptually difficult, only technically difficult”. Accordingly, and because those features that are allowed for by the general formulation but assumed away in examples where Condition (AL) is simply a representation are incidental to the main points of this paper, we have deliberately chosen illustrative examples satisfying $\sqrt{n}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^n g_0(z_i, \hat{\gamma}_n)$.

4.2. Asymptotic Separability. In many cases Condition (SE) is too strong when the goal is to prove (4) with $\mathcal{B}_n \neq o(n^{-1/2})$. This shortcoming is not shared by Condition (AS), which often turns out to be verifiable under assumptions compatible with $\beta_n \neq o(n^{-1/2})$ in Condition (AN_0). A dramatic illustration of this phenomenon is provided by Example 1.

Example 1 (continued). Suppose the kernel is of order P and satisfies standard conditions, and suppose the bandwidth satisfies $nh_n^{2P} \to 0$ and $nh_n^d \to \infty$. Then (4) holds with $\mathcal{B}_n = K(0)/(nh_n^d)$ and $\Sigma = 4\mathbb{V}[f_0(x)]$ provided f_0 is sufficiently smooth. (For details, see Section 9.)

Because $\sqrt{n}\mathcal{B}_n = K(0)/\sqrt{nh_n^{2d}}$, the condition $nh_n^d \to \infty$ is weak enough to permit $\mathcal{B}_n \neq o(n^{-1/2})$. On the other hand, (4) reduces to (1) when imposing conditions requiring $nh_n^{2d} \to \infty$, so it is necessary to guard against this when the goal is to obtain a result of the form (4) with $\mathcal{B}_n \neq o(n^{-1/2})$.

Turning first to Condition (SE) and setting $\bar{g}_0 = g_0$, the first part of that condition is automatically satisfied and the second part becomes

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\hat{f}_n(x_i) - f_n(x_i) - f_0(x_i) + \theta_0] = o_p(1), \qquad f_n(\cdot) = \mathbb{E}\hat{f}_n(\cdot).$$

It follows from a direct calculation that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\hat{f}_n(x_i) - f_n(x_i) - f_0(x_i) + \theta_0] = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o_p(1),$$

so Condition (SE) requires $nh_n^{2d} \to \infty$ and is therefore too strong for our purposes.

On the other hand, setting $\bar{g}_n = g_n$ the first part of Condition (AS) is automatically satisfied and the second part becomes

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\hat{f}_{n,i}(x_i) - 2f_n(x_i) + \theta_n] = o_p(1), \qquad \theta_n = \int_{\mathbb{R}^d} f_n(x) f_0(x) dx,$$

where $\hat{f}_{n,i}(x) = (n-1)^{-1} \sum_{j=1, j\neq i}^{n} \mathcal{K}_n(x - x_j)$. A simple variance calculation now shows that Condition (AS) is satisfied if $nh_n^d \to \infty$. ■

To interpret the bandwidth requirements $nh_n^{2d} \to \infty$ and $nh_n^d \to \infty$ implied by Conditions (SE) and (AS) in this example it is helpful to recall that the (pointwise) rate of convergence of $\hat{f}_n$ is $\sqrt{nh_n^d}$; that is, $\hat{f}_n(x) - f_n(x) = O_p(1/\sqrt{nh_n^d})$ for any $x \in \mathbb{R}^d$. The conditions $nh_n^{2d} \to \infty$ and $nh_n^d \to \infty$ therefore correspond loosely to the requirements of $n^{1/4}$-consistency and consistency, respectively, on the part of the nonparametric ingredient $\hat{f}_n$.

The fact that any rate of convergence (on the part of the nonparametric ingredient $\hat{f}_n$) will suffice when verifying Condition (AS) is largely due to the linearity of $\hat{\theta}_n$ with respect to $\hat{f}_n$. This feature is clearly special, but in three important respects Example 1 turns out to be representative. First, results of the form (4) with $\mathcal{B}_n \neq o(n^{-1/2})$ obtain only when bandwidth conditions resulting in convergence rates slower than $n^{1/4}$ are entertained on the part of nonparametric ingredients. Second, bandwidth conditions needed for stochastic equicontinuity (i.e., Condition (SE)) generally imply $n^{1/4}$-consistency on the part of nonparametric ingredients. Avoiding reliance on Condition (SE) therefore turns out to be crucial in order to achieve the goals of this paper. Third, as demonstrated by the examples that follow the qualitatively conclusion that Condition (AS) is compatible with convergence rates slower than $n^{1/4}$ on the part of nonparametric ingredients is true quite generally. Condition (AS) therefore emerges as an attractive alternative to Condition (SE).

Among estimators not satisfying (7) we are aware of only two types that provide exceptions to the rule that bandwidth conditions needed for Condition (SE) imply $n^{1/4}$-consistency on the part of nonparametric ingredients. These exceptions are those illustrated by Hausman and Newey's (1995) consumer surplus estimator and the "leave in" version of Powell, Stock, and Stoker's (1989) estimator, respectively. Hausman and Newey's (1995) consumer surplus estimator (also discussed in Newey and McFadden (1994)) is one for which $g_0(z, \gamma)$ is additively separable between z and

γ.[6] In such special cases "separability" between z and γ is of course automatic and, indeed, both parts of Condition (SE) are satisfied (without any $o_p(1)$ terms) when $g_0 = \bar{g}_0$. In the case of the "leave in" version of Powell et al.'s (1989) estimator, on the other hand, validity of Condition (SE) under weak bandwidth conditions would appear to be due to the fact that Condition (AS) holds and the fact that g_0 and g_n coincide (apart from a non-important factor of proportionality).[7]

While seemingly peculiar, the feature which makes the "leave in" version of Powell et al.'s (1989) estimator satisfy Condition (SE) turns out to be exploitable quite generally. To be specific, essentially because it involves $\hat{\gamma}_{n,i}$ rather than $\hat{\gamma}_n$, the "leave-one-out" version of Condition (SE) given by Condition (AS) turns to be verifiable under remarkably weak bandwidth conditions.

As pointed out by Newey and McFadden (1994) and illustrated by Example 1, verification (or otherwise) of the second part of Condition (SE) involves nothing more than second-order U-statistic calculations when $\bar{g}_0(z,\gamma)$ is linear in γ. Example 1 also shows that very similar calculations can be used to verify the second part of Condition (AS) when $\bar{g}_n(z,\gamma)$ is linear in γ.

To accommodate slower-than-$n^{1/4}$ rates of convergence while achieving nontrivial generality it is useful to have results that cover nonlinear-in-γ functions. It turns out that in many nonlinear-in-γ cases a variation on Newey's (1994) approach can be used to verify Condition (AS) for a $\bar{g}_n$ which is quadratic (rather than linear, as in Newey (1994)) in its second argument. For such a $\bar{g}_n$, the second part of Condition (AS) can usually by verified by means of the following lemma.

Lemma 4. *Suppose that $g_{n,\gamma}(z)[\cdot]$ is linear, $g_{n,\gamma\gamma}(z)[\cdot,\cdot]$ is bilinear, and that:*

$$\mathbb{V}(g_{n,\gamma}(z_1)[\kappa_{n,2}]) = o(n),$$

[6] Indeed, their $g_0(z,\gamma)$ does not depend on z.

[7] As pointed out by footnote 6 of Powell et al. (1989), $g_0 = (1 - n^{-1})g_n$ because symmetric kernels satisfy $K'(0) = 0$.

$$\mathbb{V}[\mathbb{E}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,2}]|z_1)] = o(n^2), \qquad \mathbb{V}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,2}]) = o(n^3),$$

$$\mathbb{V}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,3}]) = o(n^2),$$

where $\kappa_{n,j}(z,\theta) = w(z_j,\theta)\mathcal{K}_n[x(z,\theta) - x(z_j,\theta)] - \gamma_n(z,\theta)$. *Then*

$$\bar{g}_n(z,\gamma) = g_n(z,\gamma_n) + g_{n,\gamma}(z)[\gamma - \gamma_n] + \frac{1}{2}g_{n,\gamma\gamma}(z)[\gamma - \gamma_n, \gamma - \gamma_n],$$

satisfies the second part of Condition (AS).

The ease with which the first part of Condition (AS) can be verified with a quadratic $\bar{g}_n$ depends on the smoothness of $g_n(z,\gamma)$ with respect to γ. One widely applicable sufficient condition, similar in spirit to Newey (1994, Condition 5.1), is given by the following lemma, in which $\|\cdot\|_{\Gamma_0}$ denotes a norm on a function space to which $\hat{\gamma}_{n,i}(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0)$ belong (with probability approaching one).

Lemma 5. *Suppose that* $\max_{1\le i\le n}\|\hat{\gamma}_{n,i}(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0)\|_{\Gamma_0} = o_p(n^{-1/6})$ *and that for all* γ *with* $\|\gamma(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0)\|_{\Gamma_0}$ *small enough,*

$$\|g_n(z,\gamma) - \bar{g}_n(z,\gamma)\| \le b(z)\|\gamma(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0)\|^3_{\Gamma_0},$$

where $\bar{g}_n$ *is as in Lemma 4 and* $\mathbb{E}b(z) < \infty$. *Then the first part of Condition (AS) is satisfied.*

The conditions of the lemma are satisfied in Example 2, although slightly better sufficient conditions for Condition (AS) can be obtained by exploiting a special feature of the estimator considered in that example.

Example 2 (continued). A quadratic approximation to $g_n(z,f)$ is given by

$$\bar{g}_n(z,f) = g_n(z,f_n) + g_{n,f}(z)[f - f_n] + \frac{1}{2}g_{n,ff}(z)[f - f_n, f - f_n],$$

where, defining $f_n^+(x) = n^{-1}\mathcal{K}_n(0) + (1 - n^{-1})f_n(x)$,

$$\begin{aligned} g_{n,f}(z)[\kappa] &= -(1-n^{-1})\frac{y\omega(x)}{f_n^+(x)}\left[\frac{\partial}{\partial x}\kappa(x) - \frac{\partial f_n^+(x)/\partial x}{f_n^+(x)}\kappa(x)\right], \\ g_{n,ff}(z)[\kappa,\lambda] &= (1-n^{-1})^2\frac{y\omega(x)}{f_n^+(x)^2}\left[\lambda(x)\frac{\partial}{\partial x}\kappa(x) + \kappa(x)\frac{\partial}{\partial x}\lambda(x) - 2\frac{\partial f_n^+(x)/\partial x}{f_n^+(x)}\kappa(x)\lambda(x)\right]. \end{aligned}$$

If $h_n \to 0$ and if $nh_n^{\frac{3}{2}d+3}/(\log n)^{3/2} \to \infty$, then the conditions of Lemma 5 are satisfied. Proceeding as in Cattaneo et al. (2013), the second bandwidth condition can in fact be relaxed to $nh_n^{\frac{3}{2}d+1}/(\log n)^{3/2} \to \infty$. This bandwidth condition is also sufficient for the second part of Condition (AS), as the assumptions of Lemma 4 are satisfied whenever $nh_n^{d+2} \to \infty$. ■

In Example 1, Condition (AS) could be verified without imposing conditions on the rate of convergence of $\hat{f}_n$. In contrast, the condition $nh_n^{\frac{3}{2}d+1}/(\log n)^{3/2} \to \infty$ imposed in Example 2 does impose a convergence rate condition on $\hat{f}_n$, corresponding roughly to an $n^{1/6}$-consistency requirement on $\hat{f}_n$ (and its derivative). While almost certainly not the weakest possible, this bandwidth condition is attractive from the perspective of this paper as it is weak enough to permit failure of $n^{1/4}$-consistency (which once again turns out to be crucial) yet strong enough to enable us to verify Condition (AS) with a fairly modest amount of effort.

Even in cases where $g_n(z,\cdot)$ is not smooth, Lemma 5 might be useful when attempting to verify that the first part of Condition (AS) is satisfied by a quadratic $\bar{g}_n$. Example 3 provides an illustration.

Example 3 (continued). Let $F_{y|x}(\cdot|x)$ denote the conditional cdf of y given x, let $f_{y|x}(\cdot|x)$ and $\dot{f}_{y|x}(\cdot|x)$ denote its first and second derivatives. By first conditioning on x and then proceeding as in the smooth case we are led to consider

$$\bar{g}_n(x,f) = \breve{g}_n(x,f_n) + \breve{g}_{n,f}(x)[f - f_n] + \frac{1}{2}\breve{g}_{n,ff}(x)[f - f_n, f - f_n],$$

where $\breve{g}_n(x,f_n) = -F_{y|x}[f_n^+(x)|x]$ and

$$\begin{aligned}\breve{g}_{n,f}(x)[\kappa] &= -(1-n^{-1})f_{y|x}[f_n^+(x)|x]\kappa(x),\\ \breve{g}_{n,ff}(x)[\kappa,\lambda] &= -(1-n^{-1})^2\dot{f}_{y|x}[f_n^+(x)|x]\kappa(x)\lambda(x).\end{aligned}$$

If $h_n \to 0$ and if $nh_n^{\frac{3}{2}d}/(\log n)^{3/2} \to \infty$, then Lemma 5 can be used to show that the first part of Condition (AS) is satisfied. The same bandwidth conditions are sufficient for the second part of Condition (AS), as the assumptions of Lemma 4 are satisfied whenever $nh_n^d \to \infty$. ■

In perfect analogy with Example 2, also in this nonsmooth case Condition (AS) is seen hold under bandwidth conditions corresponding roughly to an $n^{1/6}$-consistency requirement on $\hat{f}_n$. Once again, the main thing to notice is the fact that Condition (AS) permits failure of $n^{1/4}$-consistency.

Remarks. (i) The second part of Condition (SE) can be verified by exhibiting a sequence $\Gamma_{n,0}$ of function classes satisfying $\lim_{n\to\infty}\mathbb{P}[\hat{\gamma}_n(\cdot,\theta_0)\in\Gamma_{n,0}] = 1$ and

$$\sup_{\gamma(\cdot,\theta_0)\in\Gamma_{n,0}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_0(z_i,\gamma) - \bar{G}_0(\gamma) - \bar{g}_0(z_i,\gamma_0) + \bar{G}_0(\gamma_0)\right\| \to_p 0,$$

where empirical process results (e.g., maximal inequalities) can be used to formulate primitive sufficient conditions for the latter. This approach, taken by Andrews (1994b, Condition (3.36)), Chen et al. (2003, Conditions (2.4) and (2.5′)), and many others, does not seem applicable when the goal is to formulate primitive sufficient conditions for the second part of Condition (AS). To be specific, the dependence of $\hat{\gamma}_{n,i}$ on i in the second part of Condition (AS) implies that the desired convergence in probability result cannot be deduced with the help of a result of the form

$$\sup_{\gamma(\cdot,\theta_0)\in\Gamma_{n,0}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\bar{g}_n(z_i,\gamma) - \bar{G}_n(\gamma) - \bar{g}_n(z_i,\gamma_n) + \bar{G}_n(\gamma_n)) \right\| \to_p 0.$$

(ii) For Example 3 arguments similar to those used in our verification of Condition (AS) can be used to show that Condition (SE) holds when $nh_n^{2d} \to \infty$ (which corresponds roughly to an $n^{1/4}$-consistency requirement on $\hat{f}_n$). This represents a considerable improvement over the condition $nh_n^{3d} \to \infty$ (which corresponds roughly to an $n^{1/3}$-consistency requirement on $\hat{f}_n$) apparently required when using empirical process methods to verify Condition (SE) by first exhibiting a function space $\mathcal{H}$ satisfying (3.3) of Chen et al. (2003)'s Theorem 3 (as is done by Chen et al. (2003, p. 1600)) and then formulating bandwidth conditions under which $\hat{f}_n$ ($\hat{h}$ in their notation) belongs to $\mathcal{H}$ with probability approaching one. ■

4.3. Asymptotic Normality. Suppose $\bar{g}_n$ is as in Lemmas 4 and 5, in which case we have

$$\bar{G}_n(\gamma) = G_n(\gamma_n) + G_{n,\gamma}[\gamma - \gamma_n] + \frac{1}{2} G_{n,\gamma\gamma}[\gamma - \gamma_n, \gamma - \gamma_n],$$

where $G_{n,\gamma}[\cdot] = \mathbb{E} g_{n,\gamma}(z)[\cdot]$ and $G_{n,\gamma\gamma}[\cdot] = \mathbb{E} g_{n,\gamma\gamma}(z)[\cdot]$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g_n(z_i,\gamma_n) + \bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_n(z_i) + \sqrt{n} B_n,$$

where, defining $\delta_n(z) = G_{n,\gamma}[w(z,\theta_0)\mathcal{K}_n(x(\cdot,\theta_0) - x(z,\theta_0)) - \gamma_n(\cdot,\theta_0)]$,

$$\psi_n(z) = g_n(z, \gamma_n) - \mathbb{E}g_n(z, \gamma_n) + \delta_n(z)$$

and

$$B_n = \mathbb{E}g_n(z, \gamma_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} G_{n,\gamma\gamma}[\hat{\gamma}_{n,i} - \gamma_n, \hat{\gamma}_{n,i} - \gamma_n].$$

The assumptions of the following lemma are usually straightforward to verify.

Lemma 6. *Suppose that*

$$\mathbb{E}[\|\psi_n(z) - \psi(z)\|^2] \to 0 \tag{9}$$

for some function ψ *and that:*

$$\mathbb{V}(G_{n,\gamma\gamma}[\kappa_{n,1}, \kappa_{n,1}]) = o(n^2), \qquad \mathbb{V}(G_{n,\gamma\gamma}[\kappa_{n,1}, \kappa_{n,2}]) = o(n). \tag{10}$$

Then Condition (AN) holds with $\Omega = \mathbb{E}[\psi(z)\psi(z)']$ *and* β_n *any sequence satisfying* $\beta_n = \mathbb{E}B_n + o(n^{-1/2})$.

Knowing the functional form of ψ and β_n is unnecessary for the purposes of verifying Condition (AN) and/or applying the bootstrap, but since ψ and β_n may nevertheless be objects of theoretical interest we comment briefly on them before verifying Condition (AN) for the examples. The function ψ in (9) will typically be given by $\psi(z) = g_0(z, \gamma_0) + \delta_0(z)$, where $\delta_0(z)$ is the "correction term" discussed by Newey (1994). Regarding the "bias" sequence β_n, we have the decomposition $\mathbb{E}B_n = \beta_n^S + \beta_n^{LI} + \beta_n^{NL}$, where

$$\beta_n^S = \mathbb{E}g_0(z, \gamma_n), \qquad \beta_n^{LI} = \mathbb{E}[g_n(z, \gamma_n) - g_0(z, \gamma_n)], \qquad \beta_n^{NL} = \frac{1}{2n}\mathbb{E}G_{n,\gamma\gamma}[\kappa_{n,1}, \kappa_{n,1}].$$

Here, β_n^S is a familiar (smoothing) bias term, which can typically be ignored because $\beta_n^S = o(n^{-1/2})$ under standard smoothness, kernel, and bandwidth conditions.[8] The term β_n^{LI} is a "leave in" bias term (in the terminology of Cattaneo et al. (2013)) whose presence can be attributed to a failure of stochastic equicontinuity. This term is nonnegligible in all of our examples. The final term, β_n^{NL}, is what Cattaneo et al. (2013) refer to as a "nonlinearity" bias term. This term is zero in Example 1 (where $\hat{\theta}_n$ is a linear functional of $\hat{f}_n$), but is nonnegligible in Examples 2 and 3 (where $\hat{\theta}_n$ is a nonlinear functional of $\hat{f}_n$). To summarize, we can typically set $\beta_n = \beta_n^{LI} + \beta_n^{NL}$.

Example 1 (continued). If $h_n \to 0$, then (9) holds with $\psi(x) = 2[f_0(x) - \theta_0]$. Also, B_n is nonrandom and satisfies $B_n = K(0)/(nh_n^d) + O(h_n^P + n^{-1})$. As a consequence, Condition (AN) is satisfied with $\Sigma = 4\mathbb{V}[f_0(x)]$ and $\beta_n = K(0)/(nh_n^d)$ provided $nh_n^{2P} \to 0$.

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then the conditions of Theorem 2 are satisfied and (4) holds with $\mathcal{B}_n = O(1/(nh_n^d)) \neq o\left(n^{-1/2}\right)$. ■

Example 2 (continued). If $h_n \to 0$ and if $nh_n^d \to \infty$, then (9) holds with $\psi(z) = g_0(z, f_0) + \delta_0(z)$, where

$$\delta_0(z) = \omega(x)\frac{\partial}{\partial x}r(x) + r(x)\frac{\partial}{\partial x}\omega(x) + r(x)\omega(x)\frac{\partial f_0(x)/\partial x}{f_0(x)}.$$

Also, (10) is satisfied if $h_n \to 0$ and if $nh_n^{d+2} \to \infty$. As a consequence, Condition (AN) is satisfied with $\Sigma = \mathbb{E}[\psi(z)\psi(z)']$ if $h_n \to 0$ and if $nh_n^{d+2} \to \infty$.

If also $nh_n^{2P} \to 0$, then $\beta_n^S = O(h_n^P) = o(n^{-1/2})$ and

[8] To be specific, we typically have $\beta_n^S = O(h_n^P)$, where P is the order of the kernel. In this case, β_n^S can be ignored if $nh_n^{2P} \to 0$.

$$
\begin{aligned}
\beta_n^{LI} &= \frac{1}{nh_n^d} K(0) \int_{\mathbb{R}^d} r(x)\omega(x) \frac{\partial f_0(x)/\partial x}{f_0(x)} dx + o(n^{-1/2}), \\
\beta_n^{NL} &= \frac{1}{nh_n^{d+1}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_0(x)} K(r) \frac{\partial}{\partial r} K(r) f_0(x - rh_n) dx dr \\
&\quad - \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_0(x)} \frac{\partial f_0(x)/\partial x}{f_0(x)} K(r)^2 f_0(x - rh_n) dx dr + o(n^{-1/2}),
\end{aligned}
$$

where it follows from Cattaneo et al. (2013, Lemma 1) that β_n^{NL} admits a polynomial-in-h_n expansion of the form

$$
\beta_n^{NL} = \frac{1}{nh_n^d} [\mathtt{B}_0 + \mathtt{B}_1 h_n^2 + \mathtt{B}_2 h_n^4 + \ldots],
$$

the constants $\mathtt{B}_0, \mathtt{B}_1, \mathtt{B}_2, \ldots$ being functionals of K and the data generating process. (Symmetry of the kernel implies that β_n^{NL} is of order $1/(nh_n^d)$ and that the polynomial expansion of $nh_n^d \beta_n^{NL}$ involves only even powers of h_n.)

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^{\frac{3}{2}d+1}/(\log n)^{3/2} \to \infty$, then the conditions of Theorem 2 are satisfied and (4) holds with $\mathcal{B}_n = O(1/(nh_n^d)) \neq o\left(n^{-1/2}\right)$. ■

Example 3 (continued). If $h_n \to 0$ and if $nh_n^d \to \infty$, then (9) holds with $\psi(z) = g_0(z, f_0) + \delta_0(x)$, where

$$
\delta_0(x) = -f_{y|x}[f_0(x)|x] f_0(x) + \int_{\mathbb{R}^d} f_{y|x}[f_0(x)|x] f_0(x)^2 dx.
$$

Also, (10) is satisfied if $h_n \to 0$ and if $nh_n^d \to \infty$. As a consequence, Condition (AN) is satisfied with $\Sigma = \mathbb{E}[\psi(z)\psi(z)']$ if $h_n \to 0$ and if $nh_n^d \to \infty$.

If also $nh_n^{2P} \to 0$, then $\beta_n^S = O(h_n^P) = o(n^{-1/2})$ and

$$\begin{aligned}
\beta_n^{LI} &= -\frac{1}{nh_n^d} K(0) \int_{\mathbb{R}^d} f_{y|x}[f_0(x)|x] f_0(x) dx + o(n^{-1/2}), \\
\beta_n^{NL} &= -\frac{1}{nh_n^d} \frac{1}{2} \int_{\mathbb{R}^d} \dot{f}_{y|x}[f_0(x)|x] K(r)^2 f_0(x) f_0(x - rh_n) dx dr + o(n^{-1/2}),
\end{aligned}$$

where β_n^{NL} can be shown to admit a polynomial-in-h_n expansion of the form

$$\beta_n^{NL} = \frac{1}{nh_n^d} [\mathtt{B}_0 + \mathtt{B}_1 h_n^2 + \mathtt{B}_2 h_n^4 + \ldots],$$

the constants $\mathtt{B}_0, \mathtt{B}_1, \mathtt{B}_2, \ldots$ being functionals of K and the data generating process.

In summary, if $nh_n^{2P} \to 0$ and if $nh_n^{\frac{3}{2}d}/(\log n)^{3/2} \to \infty$, then the conditions of Theorem 2 are satisfied and (4) holds with $\mathcal{B}_n = O(1/(nh_n^d)) \neq o\left(n^{-1/2}\right)$. ■

The findings for Examples 1 and 3 are in qualitative agreement with the those for Example 2. For the model studied in Example 2, Cattaneo et al. (2013, Section 3.3) went slightly further and used a bias expansion to develop a bandwidth selection method. The bandwidths chosen by that procedure are of order $n^{-1/(P+d)}$ and were found by Cattaneo et al. (2013) to perform well. Although it is beyond the scope of this paper to do so, it seems plausible analogous bandwidth selection procedures can be developed for Examples 1 and 3 (and perhaps even in fairly complete generality).

5. Bootstrap Consistency

One consequence of replacing (1) with (4) is that the statistics $\sqrt{n}(\hat{\theta}_n - \theta_0)$ cease to be tight. Proving bootstrap consistency without existence of limiting distributions (or even tightness) can be difficult in general (e.g., Radulovic (1998)), but thankfully the present setting has enough structure to enable us to give a simple characterization of bootstrap consistency. Indeed, in light of (5) the following condition is (necessary and) sufficient for (3) under (4) :

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n - \mathcal{B}_n) \rightsquigarrow_p \mathcal{N}(0, \Sigma), \tag{11}$$

where $\rightsquigarrow_p$ denotes weak convergence in probability.

This characterization is very useful for our purposes because it turns out that in many cases verification of (3) and (4) can proceed by first proving (4) and then, by imitating that proof, demonstrating (11) . For example, the next result is a bootstrap analogue of Theorem 2. To state it, let $\hat{\theta}_n^*$ be an (approximate) minimizer of

$$\hat{G}_n^*(\theta, \hat{\gamma}_n^*)' W \hat{G}_n^*(\theta, \hat{\gamma}_n^*), \qquad \hat{G}_n^*(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} g[z_i^*, \theta, \gamma(\cdot, \theta)],$$

where

$$\hat{\gamma}_n^*(z, \theta) = \frac{1}{n} \sum_{j=1}^{n} w(z_j^*, \theta) \mathcal{K}_n[x(z, \theta) - x(z_j^*, \theta)]$$

and $z_1^*, \ldots, z_n^*$ is a random sample with replacement from $z_1, \ldots, z_n$. Also, let

$$\hat{\gamma}_{n,i}^*(z, \theta) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} w(z_j^*, \theta) \mathcal{K}_n[x(z, \theta) - x(z_j^*, \theta)] \qquad (i = 1, \ldots, n).$$

Theorem 7. *Suppose the assumptions of Theorem 2 are satisfied and that:*
(AL) for the same $\mathcal{J}$ as in (AL),*

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \mathcal{J} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_n^*(z_i^*, \hat{\gamma}_{n,i}^*) + o_p(1),$$

where $g_n^*(z, \gamma) = g[z, \hat{\theta}_n, n^{-1} w(z, \hat{\theta}_n) \mathcal{K}_n[x(\cdot, \hat{\theta}_n) - x(z, \hat{\theta}_n)] + (1 - n^{-1}) \gamma(\cdot, \hat{\theta}_n)]$;

(AS) for some function* $\bar{g}_n^*$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_{n,i}^*)-\bar{g}_n^*(z_i^*,\hat{\gamma}_{n,i}^*)-g_n^*(z_i^*,\hat{\gamma}_n)+\bar{g}_n^*(z_i^*,\hat{\gamma}_n)]\to_p 0,$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_n^*(z_i^*,\hat{\gamma}_{n,i}^*)-\bar{G}_n^*(\hat{\gamma}_{n,i}^*)-\bar{g}_n^*(z_i^*,\hat{\gamma}_n)+\bar{G}_n^*(\hat{\gamma}_n)]\to_p 0,$$

where $\bar{G}_n^*(\gamma)=n^{-1}\sum_{i=1}^{n}\bar{g}_n^*(z_i,\gamma)$;

(AN) for the same* Ω *and* β_n *as in (AL),*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_n)+\bar{G}_n^*(\hat{\gamma}_{n,i}^*)-\bar{G}_n^*(\hat{\gamma}_n)-\beta_n]\rightsquigarrow_p \mathcal{N}(0,\Omega).$$

Then (3) *holds.*

The conditions of Theorem 7 are natural bootstrap analogs of the corresponding conditions of Theorem 2 not only in appearance but also in the sense that they can usually be verified by mimicking the verification of the corresponding conditions of Theorem 2.

Like (AL), Condition (AL*) is simply a representation in many important cases, including Examples 1-3. More generally, Condition (AL*) can often be verified with the help of a bootstrap analog of Lemma 3 given in Section 7.2.

Turning next to Conditions (AS*) and (AN*), suppose $\bar{g}_n^*$ satisfies

$$\bar{g}_n^*(z,\gamma)=g_n^*(z,\hat{\gamma}_n)+g_{n,\gamma}^*(z)[\gamma-\hat{\gamma}_n]+\frac{1}{2}g_{n,\gamma\gamma}^*(z)[\gamma-\hat{\gamma}_n,\gamma-\hat{\gamma}_n],$$

where $g_{n,\gamma}^*(z)[\cdot]$ is linear and $g_{n,\gamma\gamma}^*(z)[\cdot,\cdot]$ is bilinear. In perfect analogy with Lemma 4, the second part of Condition (AS*) is satisfied provided the following conditions hold:

$$\mathbb{V}^*(g_{n,\gamma}^*(z_1^*)[\kappa_{n,2}^*])=o_p(n),$$

$$\mathbb{V}^*[\mathbb{E}^*(g^*_{n,\gamma\gamma}(z^*_1)[\kappa^*_{n,2},\kappa^*_{n,2}]|z^*_1)] = o_p(n^2), \qquad \mathbb{V}^*(g^*_{n,\gamma\gamma}(z^*_1)[\kappa^*_{n,2},\kappa^*_{n,2}]) = o_p(n^3),$$
$$\mathbb{V}^*(g^*_{n,\gamma\gamma}(z^*_1)[\kappa^*_{n,2},\kappa^*_{n,3}]) = o_p(n^2),$$

where $\kappa^*_{n,j}(z,\theta) = w(z^*_j,\theta)\mathcal{K}_n[x(z,\theta)-x(z^*_j,\theta)] - \hat{\gamma}_n(z,\theta)$, $\mathbb{E}^*(\cdot) = \mathbb{E}(\cdot|z_1,\ldots,z_n)$, and $\mathbb{V}^*(\cdot) = \mathbb{V}(\cdot|z_1,\ldots,z_n)$.

As in the case of the first part of Condition (AS), the ease with (and methods by) which the first part of Condition (AS*) can be verified with a "quadratic" $\bar{g}^*_n$ depends on the specifics of the example, but verification by imitation is usually straightforward once verification of the first part of Condition (AS) has been achieved.

Finally, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [g^*_n(z_i,\hat{\gamma}_n) + \bar{G}^*_n(\hat{\gamma}^*_{n,i}) - \bar{G}^*_n(\hat{\gamma}_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n [\psi^*_n(z^*_i) + B^*_n],$$

where, defining $\delta^*_n(z) = G^*_{n,\gamma}[w(z,\hat{\theta}_n)\mathcal{K}_n(x(\cdot,\hat{\theta}_n) - x(z,\hat{\theta}_n)) - \hat{\gamma}_n(\cdot,\hat{\theta}_n)]$,

$$\psi^*_n(z) = g^*_n(z,\hat{\gamma}_n) - \frac{1}{n}\sum_{i=1}^n g^*_n(z_i,\hat{\gamma}_n) + \delta^*_n(z)$$

and

$$B^*_n = \frac{1}{n}\sum_{i=1}^n g^*_n(z_i,\hat{\gamma}_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^n G^*_{n,\gamma\gamma}[\hat{\gamma}^*_{n,i} - \hat{\gamma}_n, \hat{\gamma}^*_{n,i} - \hat{\gamma}_n].$$

If Condition (AN) holds, then so does Condition (AN*) if

$$\frac{1}{n}\sum_{i=1}^n \|\psi^*_n(z_i) - \psi_n(z_i)\|^2 \to_p 0$$

and if

$$\mathbb{V}^*(G^*_{n,\gamma\gamma}[\kappa^*_{n,1},\kappa^*_{n,1}]) = o_p(n^2), \qquad \mathbb{V}^*(G^*_{n,\gamma\gamma}[\kappa^*_{n,1},\kappa^*_{n,2}]) = o_p(n),$$
$$\mathbb{E}^*(B^*_n) = \beta_n + o_p(n^{-1/2}).$$

These conditions can usually be verified using straightforward, but possibly tedious, moment calculations for U-statistics.

In our examples it turns out that the conditions of Theorem 7 can be verified under the same bandwidth conditions as those used when verifying the conditions of Theorem 2. As a consequence, assuming for specificity that the bandwidth is of the form $h_n = Cn^{-1/\eta}$ (where $C > 0$ and $\eta > 0$ are user-chosen constants) the examples allow us to draw the following conclusions regarding the "robustness" of inference procedures with respect to bandwidth choice. Asymptotic validity of (the Ibragimov and Müller (2010) procedure and) the confidence interval CI_n motivated by the distributional approximation $\sqrt{n}(\hat{\theta}_n - \theta_0)\dot{\sim}\mathcal{N}(0,\hat{\Sigma}_n)$ requires $\eta \in (2d, 2P)$, while asymptotic validity of subsampling-based confidence intervals requires $\eta \in [2d, 2P)$. On the other hand, there is an example-specific constant $\eta_* < 2d$ such that the bootstrap-based confidence interval CI^*_n is asymptotically valid whenever $\eta \in (\eta_*, 2P)$. The constant η_* equals d in Example 1 and it follows from the results presented above that η_* is no greater than $\frac{3}{2}d + 1$ and $\frac{3}{2}d$ in Examples 2 and 3, respectively.[9] In other words, the range of η-values for which bootstrap-based inference procedures enjoy asymptotic validity is wider than the range for which its main rivals are valid. As a result, there

[9]The results of Cattaneo, Crump, and Jansson (2014) suggest that $\eta_* = d$ cannot be improved in Example 1. In contrast, for Examples 2 and 3 we conjecture that even smaller values of η_* can be obtained with some additional effort.

is a formal sense in which bootstrap-based inference procedures are more "robust" with respect to bandwidth choice than their main rivals.

Remarks. (i) We deliberately study only the simplest version of the bootstrap. As in Hahn (1996) doing so is sufficient when the goal is to establish first-order asymptotic validity, but we conjecture that results analogous to Theorem 2 can be obtained for various modifications of the simple nonparametric bootstrap, including those proposed by Brown and Newey (2002) and Hall and Horowitz (1996) to handle overidentified models.

(ii) Similarly, to highlight the fact that asymptotic pivotality plays no role in our theory we use the bootstrap to approximate the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ rather than a studentized version thereof.

6. Conclusion

This paper has developed "small bandwidth" asymptotic results for a class of two-step semiparametric estimators previously studied by Newey and McFadden (1994) and others. The first main result, Theorem 2, differs from those obtained in earlier work in semiparametric two-step estimators by accommodating a non-negligible bias and a noteworthy feature of the assumptions of Theorem 2 is that reliance on a commonly employed stochastic equicontinuity condition is avoided. The second main result, Theorem 7, shows that the bootstrap provides an automatic method of correcting for the bias even when it is non-negligible.

The findings of this paper are pointwise in two distinct respects. First, the distribution of observables is held fixed when developing large sample theory. Second, the results are obtained for a fixed bandwidth sequence. It would be of interest to develop uniform versions of Theorems 2 and 7 (along the lines of Romano and Shaikh (2012) and Einmahl and Mason (2005), respectively), but doing so is beyond the scope of this paper.

Although the size of the class of estimators covered by our results is nontrivial it would be of interest to explore whether conclusions analogous to ours can be obtained for semiparametric two-step estimators whose first step involves other types of nonparametric estimators (e.g., local polynomial or sieve estimators of M-regression functions). In this paper we focus on kernel-based estimators because of their analytical tractability, but we conjecture that our results (and the methods by which they are obtained) can be extended to cover other nonparametric first-step estimators. In future work we intend to attempt to substantiate this conjecture.

References

Abadie, A., and G. W. Imbens (2006): “Large Sample Properties of Matching Estimators for Average Treatment Effects,” *Econometrica*, 74, 235–267.

——— (2008): “On the Failure of the Bootstrap for Matching Estimators,” *Econometrica*, 76, 1537–1557.

——— (2011): “Bias-Corrected Matching Estimators for Average Treatment Effects,” *Journal of Business and Economic Statistics*, 29, 1–11.

Andrews, D. W. K. (1994a): “Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity,” *Econometrica*, 62, 43–72.

——— (1994b): “Empirical Process Methods in Econometrics,” in *Handbook of Econometrics, Volume 4*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, 2247-2294.

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

Brown, B. W., and W. K. Newey (2002): “Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference,” *Journal of Business and Economic Statistics*, 20, 507–517.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2013): “Generalized Jackknife Estimators of Weighted Average Derivatives (with Discussion and Rejoinder),” *Journal of the American Statistical Association*, 108, 1243–1268.

——— (2014): “Bootstrapping Density-Weighted Average Derivatives,” Econometric Theory, forthcoming.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics, Volume 6B*, ed. by J. J. Heckman, and E. E. Leamer. New York: North Holland, 5549-5632.

Chen, X., O. Linton, and I. van Keilegom (2003): "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591–1608.

Cheng, G., and J. C. Huang (2010): "Bootstrap Consistency for General Semiparametric M-Estimation," *Annals of Statistics*, 38, 2884–2915.

Einmahl, U., and D. M. Mason (2005): "Uniform in Bandwidth Consistency of Kernel-Type Function Estimators," *Annals of Statistics*, 33, 1380–1403.

Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12, 187–197.

Hall, P., and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891–916.

Hausman, J. A., and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63, 1445–1476.

Ibragimov, R., and U. K. Müller (2010): "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468.

Ichimura, H., and O. Linton (2005): "Asymptotic Expansions for some Semiparametric Program Evaluation Estimators," in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. New York: Cambridge University Press, 149-170.

ICHIMURA, H., AND P. E. TODD (2007): "Implementing Nonparametric and Semiparametric Estimators," in *Handbook of Econometrics, Volume 6B*, ed. by J. J. Heckman, and E. E. Leamer. New York: North Holland, 5369-5468.

LINTON, O. (1995): "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63, 1079–1112.

MAMMEN, E. (1989): "Asymptotics with Increasing Dimension for Robust Regression with Applications to the Bootstrap," *Annals of Statistics*, 17, 382–400.

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

NEWEY, W. K., F. HSIEH, AND J. M. ROBINS (2004): "Twicing Kernels and a Small Bias Property of Semiparametric Estimators," *Econometrica*, 72, 947–962.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Volume 4*, ed. by R. F. Engle, and D. L. McFadden. New York: North Holland, 2111-2245.

NISHIYAMA, Y., AND P. M. ROBINSON (2000): "Edgeworth Expansions for Semiparametric Averaged Derivatives," *Econometrica*, 68, 931–979.

——— (2001): "Studentization in Edgeworth Expansions for Estimates of Semiparametric Index Models," in *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. L. Powell. New York: Cambridge University Press, 197-240.

——— (2005): "The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives," *Econometrica*, 73, 903–948.

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

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

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Radulovic, D. (1998): "Can We Bootstrap Even if CLT Fails?," *Journal of Theoretical Probability*, 11, 813–830.

Romano, J. P., and A. M. Shaikh (2012): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap," *Annals of Statistics*, 40, 2798–2822.

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

7. Appendix I: Additional Results

7.1. Orthogonality. In the case where the nonparametric estimator $\hat{\gamma}_n$ satisfies $\hat{\gamma}_n - \gamma_0 = o_p(n^{-1/4})$, sufficient conditions for the "orthogonality" condition (7) to hold have been given by Newey (1994), among others. For completeness, this subsection discusses conditions for (7) to hold also in the case where only $\hat{\gamma}_n - \gamma_0 = o_p(n^{-1/6})$ is assumed. In this case g_0 will typically admit linear and bilinear functionals $g_{0,\gamma}(z)[\cdot]$ and $g_{0,\gamma\gamma}(z)[\cdot,\cdot]$ such that the first part of Condition (SE) holds with

$$\bar{g}_0(z,\gamma) = g_0(z,\gamma_0) + g_{0,\gamma}(z)[\gamma-\gamma_0] + \frac{1}{2} g_{0,\gamma\gamma}(z)[\gamma-\gamma_0,\gamma-\gamma_0].$$

Under mild additional moment conditions, (7) then holds if

$$\mathbb{E} g_{0,\gamma}(z)[\cdot] = 0 \tag{12}$$

and if

$$\mathbb{E} g_{0,\gamma\gamma}(z)[\cdot,\cdot] = 0. \tag{13}$$

Both (12) and (13) are usually straightforward to verify (when they hold). For instance, verifying the conditions is easy when $g(z,\theta,\gamma) = A(x(z),\theta,\gamma)m(z,\theta)$, where $m(z,\theta)$ is a residual satisfying $\mathbb{E}[m(z,\theta_0)|x(z)] = 0$ and $A(x(z),\theta,\gamma)$ is a matrix of (optimal) instruments depending on θ and γ (e.g., Newey (1990)).

The condition (12) is similar to Newey (1994, Proposition 3). Under this condition, the "correction term" of Newey (1994) is zero and (12) therefore implies that the presence of the nonparametric estimator has no impact on the asymptotic variance of $\hat{\theta}_n$. Unlike (12) , (13) has no counterpart in Newey (1994), but is needed when the condition $\hat{\gamma}_n - \gamma_0 = o_p(n^{-1/4})$ is relaxed. The condition (13) implies that $\mathcal{B}_n$

in (4) can be set equal to zero even under "small bandwidth" asymptotics. As a consequence, (13) can be thought of as giving a condition under which the presence of the nonparametric estimator has no impact on the asymptotic bias of $\hat{\theta}_n$.

It is easy to give examples where (12) holds, but (13) and (7) do not. As a consequence, having a "correction term" equal to zero is only a necessary condition for (7) to hold when $\hat{\gamma}_n - \gamma_0 \neq o_p(n^{-1/4})$. In particular, also estimators having the "small bias property" discussed by Newey, Hsieh, and Robins (2004) can suffer from the bias problems highlighted in this paper.

7.2. Remarks on Lemma 3. The assumptions of Lemma 3 are comparable to those of Pakes and Pollard (1989, Theorem 3.3). As in Pakes and Pollard (1989, Theorem 3.3), one of the conditions of Lemma 3 is a consistency condition on $\hat{\theta}_n$. That condition can usually be verified by adapting Pakes and Pollard (1989, Corollary 3.2) to our setup. The condition $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$ has no counterpart in Pakes and Pollard (1989, Theorem 3.3), but is satisfied in the cases of primary interest in this paper and is therefore included as an assumption in Lemma 3. Our conditions (i) and (v) correspond exactly to the analogous conditions in Pakes and Pollard (1989, Theorem 3.3).[10]

Condition (iv) of Lemma 3 is weaker than its counterpart in Pakes and Pollard (1989, Theorem 3.3), a weakening necessitated by the fact that in the cases of main interest in this paper only an assumption of the form $\hat{G}_n(\theta_0, \hat{\gamma}_n) = o_p(n^{-1/3})$ can be justified. To compensate for this weakening, conditions (ii) and (iii) of Lemma 3 are somewhat stronger than the natural counterparts of conditions (ii) and (iii) of Pakes and Pollard (1989, Theorem 3.3). For instance, while the natural counterpart

[10] As in Pakes and Pollard (1989, pp. 1044-1046), it is possible to relax the assumption that the matrix W in (i) is nonrandom. To be specific, the conclusion of Lemma 3 remains unaffected if W is replaced with an estimator $\hat{W}_n$ satisfying $\hat{W}_n = W + o_p(n^{-1/6})$.

of condition (ii) of Pakes and Pollard (1989, Theorem 3.3) imposes only a full rank condition on $\dot{G}_0' W \dot{G}_0$ our condition (ii) also includes additional (mild) smoothness requirements on G.

Translated into our notation, the natural counterpart of condition (iii) of Pakes and Pollard (1989, Theorem 3.3) is intended to ensure that

$$\|\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\hat{\theta}_n, \gamma_0) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \gamma_0)\| = o_p(n^{-1/2}).$$

Under our condition (ii), the $\alpha = 0$ version of our condition (iii) achieves the same goal when $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$ and $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$. The $\alpha = 1/3$ version of condition (iii) has no obvious counterpart in Pakes and Pollard (1989, Theorem 3.3), but can be interpreted as a weakened version of a condition similar to (13) of Cheng and Huang (2010, Condition S1). This condition is used here to handle situations where $\theta_0 \in \mathbb{R}^k$ is "overidentified" in the sense the dimension of g exceeds k.[11]

When $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$ and $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, a seemingly more elegant way of arriving at the displayed result would be to set $\Gamma_n = \Gamma(n^{-1/6})$ and assume the following variant of Chen et al. (2003)'s Condition $(2.5')$: [12]

(SE') for every positive $\delta_n = o(n^{-1/3})$,

$$\sup_{\theta \in \Theta(\delta_n), \gamma \in \Gamma_n} \|\hat{G}_n(\theta, \gamma) - G(\theta, \gamma) - \hat{G}_n(\theta_0, \gamma_0) + G(\theta_0, \gamma_0)\| = o_p(n^{-1/2}).$$

[11] In the "just identified" case where θ_0 and g are of the same dimension, the $\alpha = 1/3$ version of condition (iii) of Lemma 3 can be dropped (and any $s > 0$ suffices when verifying the $\alpha = 0$ version of condition (iii) by means of Lemma 8).

[12] By assumption, $G(\theta_0, \gamma_0) = 0$. The purpose of retaining $G(\theta_0, \gamma_0)$ in the formulation of Condition (SE') is to facilitate comparison with condition (iii) of Lemma 3. We use the label (SE') in recognition of the fact that the condition is stronger than Condition (SE) whenever $\hat{\gamma}_n \in \Gamma_n$ (with probability approaching one).

An implication of the discussion in Section 4.2 is that Condition (SE') is violated in the examples of interest in this paper. In contrast, because the stochastic equicontinuity condition (iii) of Lemma 3 employs a formulation in which the second argument of each function inside the $\|\cdot\|$ is the same, the condition often reduces to a condition concerning the fluctuations of a process indexed by the finite-dimensional parameter θ, suggesting that it might be verifiable under weaker conditions than Condition (SE'), which concerns the fluctuations of a process which depends on the infinite-dimensional parameter γ as well. Indeed, condition (iii) of Lemma 3 is automatically satisfied (with the $o_p(n^{-1/2-\alpha/2})$ term equal to zero) in Examples 1-3.[13] More generally, the condition can often be verified by setting $s \geq 1/2$ and $\Gamma_n = \Gamma(n^{-1/6})$ and verifying the condition of the following lemma.

Lemma 8. *Let $\alpha \geq 0$ and suppose that*

$$\overline{\lim}_{n\to\infty}\overline{\lim}_{\delta\downarrow 0}\delta^{-2s}\sup_{\theta\in\aleph}\mathbb{E}[F_n(z;\delta,\theta)^2] < \infty$$

for some neighborhood $\aleph$ of θ_0 and some $s > 0$, where

$$F_n(z;\delta,\theta) = \sup_{\|\theta'-\theta\|\leq\delta,\gamma\in\Gamma_n} \|g(z,\theta',\gamma(\cdot,\theta')) - g(z,\theta,\gamma(\cdot,\theta)\|.$$

Then, for every $\alpha \geq 0$ and for every positive $\delta_n = o(n^{-\alpha})$,

$$\sup_{\theta\in\Theta(\delta_n),\gamma\in\Gamma_n} \|\hat{G}_n(\theta,\gamma) - G(\theta,\gamma) - \hat{G}_n(\theta_0,\gamma) + G(\theta_0,\gamma)\| = o_p(n^{-1/2-\alpha s}) \qquad (14)$$

and

$$\sup_{\theta\in\Theta(\delta_n),\gamma\in\Gamma_n} \|\hat{G}^*_n(\theta,\gamma) - G(\theta,\gamma) - \hat{G}^*_n(\theta_0,\gamma) + G(\theta_0,\gamma)\| = o_p(n^{-1/2-\alpha s}). \qquad (15)$$

[13] Because Condition (AL) itself is automatically satisfied in Examples 1-3 there is no need to verify it using Lemma 3. Nevertheless, given the failure of Condition (SE') in these examples it seems comforting that condition (iii) of Lemma 3 is satisfied in the examples.

Like Chen et al. (2003, Theorem 3), Lemma 8 is in the spirit of Andrews (1994b, Section 5). The lemma is obtained by bounding the $\mathcal{L}^2$ bracketing numbers of the classes

$$\mathcal{F}_n = \{g(\cdot,\theta,\gamma(\cdot,\theta)) - g(\cdot,\theta_0,\gamma(\cdot,\theta_0)) : \theta \in \Theta(\delta_n), \gamma \in \Gamma_n\}.$$

These bracketing numbers are often considerably smaller in magnitude than the bracketing numbers of classes such as $\{g(\cdot,\theta,\gamma(\cdot,\theta)) : \theta \in \Theta(\delta_n), \gamma \in \Gamma_n\}$, the latter being the classes of interest when verifying Condition (SE').

In Lemma 8, (15) is a natural bootstrap counterpart of (14). It is stated in anticipation of condition (iii*) of the following bootstrap analog of Lemma 3.

Lemma 9. *Suppose that $\hat{\theta}_n^* - \hat{\theta}_n = o_p(1)$, $\|\hat{\gamma}_n^* - \hat{\gamma}_n\|_\Gamma = o_p(n^{-1/6})$, that the assumptions of Lemma 3 are satisfied, and that:*
(i) $\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \le \inf_{\theta\in\Theta} \hat{G}_n^*(\theta, \hat{\gamma}_n^*)' W \hat{G}_n^*(\theta, \hat{\gamma}_n^*) + o_p(n^{-1})$;*
(iii) for $\alpha \in \{0, 1/3\}$ and for every positive $\delta_n = o(n^{-\alpha})$,*

$$\sup_{\theta\in\Theta(\delta_n)} \|\hat{G}_n^*(\theta,\hat{\gamma}_n^*) - G(\theta,\hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0,\hat{\gamma}_n^*) + G(\theta_0,\hat{\gamma}_n^*)\| = o_p(n^{-1/2-\alpha/2});$$

(iv) $\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) = o_p(n^{-1/3})$.*
Then Condition (AL) holds.*

7.3. Uniform Convergence Rates. Various results on uniform convergence rates for kernel estimators are used to verify the conditions of Theorems 2 and 7 in Examples 2 and 3. The results utilized are all special cases of Lemma 10 below.

Suppose that for every n, $Z_{in} = (W_{in}, X_i')'$ $(i = 1, \ldots, n)$ are *i.i.d.* copies of $Z_n = (W_n, X')'$, where $X \in \mathbb{R}^d$ is continuous with bounded density $f_X(\cdot)$. The kernel estimators we consider are of the form

$$\hat{\Psi}_n(x) = \frac{1}{n}\sum_{j=1}^{n} W_{jn}\kappa_n(x - X_j), \qquad \kappa_n(x) = \frac{1}{h_n^d}\kappa\left(\frac{x}{h_n}\right),$$

and

$$\hat{\Psi}_{n,i}(x) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} W_{jn}\kappa_n(x - X_j) \qquad (i = 1, \ldots, n),$$

where $h_n = o(1)$ is a bandwidth and $\kappa : \mathbb{R}^d \to \mathbb{R}$ is a bounded and integrable kernel-like function.

Bootstrap analogs of these estimators are also of interest. Letting $\{Z^*_{1n}, \ldots, Z^*_{nn}\}$ be a random sample with replacement from $\{Z_{1n}, \ldots, Z_{nn}\}$, define

$$\hat{\Psi}^*_n(x) = \frac{1}{n}\sum_{j=1}^{n} W^*_{jn}\kappa_n(x - X^*_j)$$

and

$$\hat{\Psi}^*_{n,i}(x) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} W^*_{jn}\kappa_n(x - X^*_j) \qquad (i = 1, \ldots, n).$$

Defining $\Psi_n(x) = \mathbb{E}\hat{\Psi}_n(x)$ the objective is to give conditions (on h_n, ρ_n, and the distribution of Z_n) under which

$$\max\nolimits_{1\leq i\leq n} |\hat{\Psi}_n(X_i) - \Psi_n(X_i)| = O_p(\rho_n), \tag{16}$$

$$\max\nolimits_{1\leq i\leq n} |\hat{\Psi}_{n,i}(X_i) - \Psi_n(X_i)| = O_p(\rho_n), \tag{17}$$

$$\max\nolimits_{1\leq j\leq n} |\hat{\Psi}^*_n(X_j) - \hat{\Psi}_n(X_j)| = O_p(\rho_n), \tag{18}$$

$$\max\nolimits_{1\leq i,j\leq n} |\hat{\Psi}^*_{n,i}(X_j) - \hat{\Psi}_n(X_j)| = O_p(\rho_n). \tag{19}$$

To give a succinct statement, let $Gam(\cdot)$ be the Gamma function and for $s > 0$, let

$$\mathcal{C}(s) = \sup_n [\mathbb{E}(|W_n|^s) + \sup_{x \in \mathbb{R}^d} \mathbb{E}(|W_n|^s | X = x) f_X(x)].$$

Lemma 10. *(a) If $\mathcal{C}(S) < \infty$ for some $S \geq 2$ and if $n^{1-1/S} h_n^d / \log n \to \infty$, then* $(16) - (19)$ *hold with* $\rho_n = \max(\sqrt{\log n}/\sqrt{nh_n^d}, \log n/(n^{1-1/S} h_n^d))$.
(b) If $\mathcal{C}(s) \leq Gam(s) H^s$ for some $H < \infty$ and every s and if $\underline{\lim}_{n\to\infty} nh_n^d/(\log n)^3 > 0$, then $(16) - (19)$ *hold with* $\rho_n = \sqrt{\log n}/\sqrt{nh_n^d}$.
(c) If $\mathcal{C}(s) \leq H^s$ for some $H < \infty$ and every s and if $\underline{\lim}_{n\to\infty} nh_n^d/\log n > 0$, then $(16) - (19)$ *hold with* $\rho_n = \sqrt{\log n}/\sqrt{nh_n^d}$.

The condition $\mathcal{C}(s) \leq H^s$ (for some $H < \infty$ and every s) is satisfied when W_n is bounded (uniformly in n), so part (c) can be used to analyze $\hat{f}_n$ and its derivative and we use this part in all of the examples. Part (b) covers certain distributions with full support (e.g., sub-Gaussian distributions), but is not used in our examples. On the other hand, the $S = 4$ version of part (a) is used to verify Condition (AN*) in Example 2.

8. Appendix II: Proofs

8.1. Proof of Lemma 1. The proof is elementary:

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_0(z_i, \hat{\gamma}_n) + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [g_0(z_i, \gamma_0) + \bar{g}_0(z_i, \hat{\gamma}_n) - \bar{g}_0(z_i, \gamma_0)] + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [g_0(z_i, \gamma_0) + \bar{G}_0(\hat{\gamma}_n) - \bar{G}_0(\gamma_0)] + o_p(1) \\
&\rightsquigarrow \mathcal{N}(0, \mathcal{J}\Omega\mathcal{J}'),
\end{aligned}
$$

where the first equality uses Condition (AL), the second and third equalities use Condition (SE), and the last line uses Condition (AN_0).

8.2. Proof of Theorem 2. The proof is elementary:

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathcal{J}\beta_n) &= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [g_n(z_i, \hat{\gamma}_{n,i}) - \beta_n] + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [g_n(z_i, \gamma_n) + \bar{g}_n(z_i, \hat{\gamma}_{n,i}) - \bar{g}_n(z_i, \gamma_n) - \beta_n] + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} [g_n(z_i, \gamma_n) + \bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n) - \beta_n] + o_p(1) \\
&\rightsquigarrow \mathcal{N}(0, \mathcal{J}\Omega\mathcal{J}'),
\end{aligned}
$$

where the first equality uses Condition (AL), the second and third equalities use Condition (AS), and the last line uses Condition (AN).

8.3. Proof of Lemma 3. Let $\dot{G}_n = \dot{G}(\theta_0, \hat{\gamma}_n)$ and define $\mathcal{J}_n = -(\dot{G}_n' W \dot{G}_n)^{-1}\dot{G}_n' W$. Using $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, (ii), and (iv),

$$(\mathcal{J}_n - \mathcal{J})\hat{G}_n(\theta_0, \hat{\gamma}_n) = o_p(n^{-1/2}).$$

It therefore suffices to show that $\hat{\theta}_n - \theta_0 - \mathcal{J}_n \hat{G}_n(\theta_0, \hat{\gamma}_n) = o_p(n^{-1/2})$.

To do so, let $L_n(\theta) = \dot{G}_n' W[\hat{G}_n(\theta_0, \hat{\gamma}_n) + \dot{G}_n(\theta - \theta_0)]$. Because

$$L_n(\hat{\theta}_n) = \dot{G}_n' W \dot{G}_n[\hat{\theta}_n - \theta_0 - \mathcal{J}_n \hat{G}_n(\theta_0, \hat{\gamma}_n)]$$

and because $\dot{G}_n' W \dot{G}_n \to_p \dot{G}_0' W \dot{G}_0 > 0$, it suffices to show that $L_n(\hat{\theta}_n) = o_p(n^{-1/2})$.

Using $\hat{\theta}_n - \theta_0 = o_p(1)$, $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, and (ii)-(iii) (with $\alpha = 0$),

$$\|G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) + \hat{G}_n(\theta_0, \hat{\gamma}_n)\| = o_p(n^{-1/2})$$

and

$$\|\dot{G}_n(\hat{\theta}_n - \theta_0) - G(\hat{\theta}_n, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\| = \|\hat{\theta}_n - \theta_0\|^2 O_p(1).$$

As a consequence, by the triangle inequality and using $\|\dot{G}_n' W\| = O(1)$,

$$\begin{aligned}
\|L_n(\hat{\theta}_n)\| &\leq \|\dot{G}_n' W\| \|\dot{G}_n(\hat{\theta}_n - \theta_0) - G(\hat{\theta}_n, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\| \\
&\quad + \|\dot{G}_n' W\| \|G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) + \hat{G}_n(\theta_0, \hat{\gamma}_n)\| \\
&\quad + \|\dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)\| \\
&= \|\hat{\theta}_n - \theta_0\|^2 O_p(1) + \|\dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)\| + o_p(n^{-1/2}),
\end{aligned}$$

so it suffices to show that $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$ and that $\dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_p(n^{-1/2})$.

Proof of $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$. Condition (ii) implies the existence of a positive constant C for which $\|\theta - \theta_0\| \leq C^{-1}\|G(\theta, \gamma_0)\|$ near θ_0. Because $\hat{\theta}_n - \theta_0 = o_p(1)$, it therefore suffices to show that $\|G(\hat{\theta}_n, \gamma_0)\| = \|\hat{\theta}_n - \theta_0\| o_p(1) + o_p(n^{-1/3})$.

Using (i) and (iv)-(v), we have

$$\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \leq \hat{G}_n(\theta_0, \hat{\gamma}_n)' W \hat{G}_n(\theta_0, \hat{\gamma}_n) + o_p(n^{-1}) = o_p(n^{-2/3}),$$

implying in particular that $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_p(n^{-1/3})$. Also, using $\hat{\theta}_n - \theta_0 = o_p(1)$, $\|\hat{\gamma}_n - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, and (ii)-(iii) (with $\alpha = 0$),

$$\|G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) + \hat{G}_n(\theta_0, \hat{\gamma}_n)\| = o_p(n^{-1/2})$$

and (with probability approaching one)

$$\|G(\hat{\theta}_n, \gamma_0) - G(\hat{\theta}_n, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\| \leq \|\hat{\theta}_n - \theta_0\| o_p(1).$$

Using these rates, the triangle inequality, and (iv),

$$\begin{aligned} \|G(\hat{\theta}_n, \gamma_0)\| &\leq \|G(\hat{\theta}_n, \gamma_0) - G(\hat{\theta}_n, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\| \\ &\quad + \|G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) + \hat{G}_n(\theta_0, \hat{\gamma}_n)\| \\ &\quad + \|\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)\| + \|\hat{G}_n(\theta_0, \hat{\gamma}_n)\| \\ &= \|\hat{\theta}_n - \theta_0\| o_p(1) + o_p(n^{-1/3}). \end{aligned}$$

Proof of $\dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_p(n^{-1/2})$. Because $\dot{G}_n' W \dot{G}_n \to_p \dot{G}_0' W \dot{G}_0 > 0$, it suffices to show that

$$\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \dot{G}_n (\dot{G}_n' W \dot{G}_n)^{-1} \dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_p(n^{-1}).$$

To do so, let $\tilde{\theta}_n = \hat{\theta}_n + \mathcal{J}_n \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)$, which satisfies $\tilde{\theta}_n - \theta_0 = o_p(n^{-1/3})$ because $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$ and $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_p(n^{-1/3})$.

By (ii) and (iii) (with $\alpha = 1/3$), we therefore have

$$R_n = \hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - \dot{G}_n(\tilde{\theta}_n - \hat{\theta}_n) = o_p(n^{-2/3}).$$

As a consequence, using (i) and (v), with probability approaching one

$$\begin{aligned}
\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) &\leq \hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n)' W \hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n) + o_p(n^{-1}) \\
&= \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\quad - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \dot{G}_n (\dot{G}_n' W \dot{G}_n)^{-1} \dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\quad + 2R_n' [W - W \dot{G}_n (\dot{G}_n' W \dot{G}_n)^{-1} \dot{G}_n' W] \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\quad + R_n' W R_n + o_p(n^{-1}),
\end{aligned}$$

which rearranges as

$$\begin{aligned}
&\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' W \dot{G}_n (\dot{G}_n' W \dot{G}_n)^{-1} \dot{G}_n' W \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
\leq\ & 2R_n' [W - W \dot{G}_n (\dot{G}_n' W \dot{G}_n)^{-1} \dot{G}_n' W] \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) + R_n' W R_n + o_p(n^{-1}) \\
=\ & o_p(n^{-1}).
\end{aligned}$$

8.4. Proof of Lemma 4. When $\bar{g}_n$ is defined as in the lemma, we have:

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_n(z_i,\hat{\gamma}_{n,i})-\bar{G}_n(\hat{\gamma}_{n,i})-\bar{g}_n(z_i,\gamma_n)+\bar{G}_n(\gamma_n)] \\
= \; & \frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(g_{n,\gamma}(z_i)[\kappa_{n,j}]-G_{n,\gamma}[\kappa_{n,j}]) \\
& +\frac{1}{2\sqrt{n}(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(g_{n,\gamma\gamma}(z_i)[\kappa_{n,j},\kappa_{n,j}]-G_{n,\gamma\gamma}[\kappa_{n,j},\kappa_{n,j}]) \\
& +\frac{1}{2\sqrt{n}(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\notin\{i,j\}}^{n}(g_{n,\gamma\gamma}(z_i)[\kappa_{n,j},\kappa_{n,k}]-G_{n,\gamma\gamma}[\kappa_{n,j},\kappa_{n,k}]),
\end{aligned}$$

where $G_{n,\gamma}[\cdot]=\mathbb{E}g_{n,\gamma}(z)[\cdot]$ and $G_{n,\gamma\gamma}[\cdot]=\mathbb{E}g_{n,\gamma\gamma}(z)[\cdot]$. By construction,

$$\mathbb{E}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_n(z_i,\hat{\gamma}_{n,i})-\bar{G}_n(\hat{\gamma}_{n,i})-\bar{g}_n(z_i,\gamma_n)+\bar{G}_n(\gamma_n)])=0.$$

The result therefore follows from Chebychev's inequality because

$$\begin{aligned}
& \mathbb{V}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\bar{g}_n(z_i,\hat{\gamma}_{n,i})-\bar{G}_n(\hat{\gamma}_{n,i})-\bar{g}_n(z_i,\gamma_n)+\bar{G}_n(\gamma_n)]) \\
= \; & \frac{1}{n}O(\mathbb{V}(g_{n,\gamma}(z_1)[\kappa_{n,2}])) \\
& +\frac{1}{n^2}O(\mathbb{V}[\mathbb{E}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,2}]|z_1)])+\frac{1}{n^3}O(\mathbb{V}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,2}])) \\
& +\frac{1}{n^2}O(\mathbb{V}(g_{n,\gamma\gamma}(z_1)[\kappa_{n,2},\kappa_{n,3}])) \\
= \; & o(n^{-1}),
\end{aligned}$$

where the first equality uses Hoeffding's theorem for U-statistics.

8.5. Proof of Lemma 5. Because $\max_{1\leq i\leq n}\|\hat{\gamma}_{n,i}(\cdot,\theta_0)-\gamma_n(\cdot,\theta_0)\|_{\Gamma_0}=o_p(n^{-1/6})$, we have (with probability approaching one)

$$
\begin{aligned}
& \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g_n(z_i,\hat{\gamma}_{n,i}) - \bar{g}_n(z_i,\hat{\gamma}_{n,i}) - g_n(z_i,\gamma_n) + \bar{g}_n(z_i,\gamma_n)] \right\| \\
\leq \;& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \| g_n(z_i,\hat{\gamma}_{n,i}) - \bar{g}_n(z_i,\hat{\gamma}_{n,i}) \| \\
\leq \;& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b(z_i) \| \hat{\gamma}_{n,i}(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0) \|_{\Gamma_0}^3 \\
\leq \;& (n^{1/6} \max\nolimits_{1\leq i\leq n} \| \hat{\gamma}_{n,i}(\cdot,\theta_0) - \gamma_n(\cdot,\theta_0) \|_{\Gamma_0})^3 \frac{1}{n} \sum_{i=1}^{n} b(z_i) \\
= \;& o_p(1),
\end{aligned}
$$

where the last equality uses $\mathbb{E}|b(z)| = \mathbb{E}b(z) < \infty$.

8.6. Proof of Lemma 6. If (9) holds and if $\mathbb{V}(B_n) = o(n^{-1})$, then Condition (AN) holds with $\Omega = \mathbb{E}[\psi(z)\psi(z)']$ and $\beta_n = \mathbb{E}B_n + o(n^{-1/2})$ because

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g_n(z_i,\gamma_n) + \bar{G}_n(\hat{\gamma}_{n,i}) - \bar{G}_n(\gamma_n) - \beta_n] \\
= \;& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_n(z_i) + \sqrt{n}(B_n - \mathbb{E}B_n) + \sqrt{n}(\mathbb{E}B_n - \beta_n) \\
= \;& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(z_i) + o_p(1) \rightsquigarrow \mathcal{N}(0,\Omega).
\end{aligned}
$$

Now,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} G_{n,\gamma\gamma}[\hat{\gamma}_{n,i}-\gamma_n,\hat{\gamma}_{n,i}-\gamma_n] &= \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} G_{n,\gamma\gamma}[\kappa_{n,j},\kappa_{n,j}] \\
&\quad + \frac{1}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\notin\{i,j\}}^{n} G_{n,\gamma\gamma}[\kappa_{n,j},\kappa_{n,k}] \\
&= \frac{n-1}{n(n-1)^2}\sum_{i=1}^{n} G_{n,\gamma\gamma}[\kappa_{n,i},\kappa_{n,i}] \\
&\quad + \frac{n-2}{n(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} G_{n,\gamma\gamma}[\kappa_{n,i},\kappa_{n,j}],
\end{aligned}$$

so it follows from Hoeffding's theorem for U-statistics that

$$\mathbb{V}(B_n) = \frac{1}{n^3}O(\mathbb{V}(G_{n,\gamma\gamma}[\kappa_{n,1},\kappa_{n,1}])) + \frac{1}{n^2}O(\mathbb{V}(G_{n,\gamma\gamma}[\kappa_{n,1},\kappa_{n,2}])),$$

implying in particular that $\mathbb{V}(B_n) = o(n^{-1})$ if (10) holds.

8.7. Proof of Theorem 7. It suffices to show that (11) holds with $\mathcal{B}_n = \mathcal{J}\beta_n$ and $\Sigma = \mathcal{J}\Omega\mathcal{J}'$. The proof of that result is elementary:

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n - \mathcal{J}\beta_n) &= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_{n,i}^*) - \beta_n] + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_n) + \bar{g}_n^*(z_i^*,\hat{\gamma}_{n,i}^*) - \bar{g}_n^*(z_i^*,\hat{\gamma}_n) - \beta_n] + o_p(1) \\
&= \mathcal{J}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{\gamma}_n) + \bar{G}_n^*(\hat{\gamma}_{n,i}^*) - \bar{G}_n^*(\hat{\gamma}_n) - \beta_n] + o_p(1) \\
&\rightsquigarrow_p \mathcal{N}(0,\mathcal{J}\Omega\mathcal{J}'),
\end{aligned}$$

where the first equality uses Condition (AL*), the second and third equalities use

Condition (AS*), and the last line uses Condition (AN*).

8.8. Proof of Lemma 8. Let $\alpha \geq 0$ be given and let $\delta_n = o(n^{-\alpha})$ be positive.

Studying one component at a time, if necessary, we may assume without loss of generality that $g(\cdot)$ is scalar.

Suppose $\mathbb{E}[F_n(z)^2] = O(1)$, where $F_n(z) = \sup_{f \in \mathcal{F}_n} |f(z)|$. By van der Vaart (1998, Lemma 19.35),

$$\mathbb{E}\left[\sup_{\theta \in \Theta(\delta_n), \gamma \in \Gamma_n} \|\hat{G}_n(\theta, \gamma) - G(\theta, \gamma) - \hat{G}_n(\theta_0, \gamma) + G(\theta_0, \gamma)\|\right] = O[n^{-1/2} J_{[\,]}(\mathcal{F}_n)],$$

where, with $N_{[\,]}(\varepsilon, \mathcal{F}_n)$ denoting the minimum number of ε-brackets in $\mathcal{L}^2$ needed to cover $\mathcal{F}_n$, $J_{[\,]}(\mathcal{F}_n) = \int_0^{\sqrt{\mathbb{E}[F_n(z)^2]}} \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}_n)} d\varepsilon$. To show (14) it therefore suffices to show that $\mathbb{E}[F_n(z)^2] = O(1)$ and that $J_{[\,]}(\mathcal{F}_n) = o(n^{-\alpha s})$. To do so, it suffices to show that $\mathbb{E}[F_n(z)^2] = o(n^{-2s\alpha})$ and that, for some $K > 0$, some $r > 0$, and for all n large enough,

$$\log N_{[\,]}(\varepsilon, \mathcal{F}_n) \leq \max(K - r \log[\varepsilon \delta_n^{-s}], 0).$$

By assumption,

$$\mathbb{E}[F_n(z)^2] = \mathbb{E}[F_n(z; \delta_n, \theta_0)^2] \leq \sup_{\theta \in \aleph} \mathbb{E}[F_n(z; \delta_n, \theta)^2] = O(\delta_n^{2s}) = o(n^{-2s\alpha}).$$

Also, if $M > \overline{\lim}_{n \to \infty} \overline{\lim}_{\delta \downarrow 0} \delta^{-2s} \sup_{\theta \in \aleph} \mathbb{E}[F_n(z; \delta, \theta)^2]$, then

$$\sup_{\theta \in \Theta(\delta_n)} \mathbb{E}[F_n(z; \delta, \theta)^2] \leq M \delta^{2s}$$

for all (n, δ) with n large enough and δ small enough. Given such a pair (n, δ), let $\{\theta_j : j = 1, \dots, N_n(\delta)\}$ be a δ-cover for $\Theta(\delta_n)$. The cover may be assumed to be chosen in such way that, for some constant C,

$$N_n(\delta) \leq \max \left[C \left(\frac{\delta_n}{\delta} \right)^k, 1 \right].$$

Then an $\varepsilon = \sqrt{4M\delta^{2s}}$-bracketing for $\mathcal{F}_n$ is given by

$$\{ \left[g^{\Delta}(\cdot, \theta_j) - F_n(\cdot; \delta, \theta_j), g^{\Delta}(\cdot, \theta_j) + F_n(\cdot; \delta, \theta_j) \right] : j = 1, \dots, N_n(\delta) \},$$

where $g^{\Delta}(\cdot, \theta_j) = g(\cdot, \theta_j, \gamma_0(\cdot, \theta_j)) - g(\cdot, \theta_0, \gamma_0(\cdot, \theta_0))$. As a consequence,

$$\log N_{[\,]}(\varepsilon, \mathcal{F}_n) \leq \log N_n \left(\frac{\varepsilon^{1/s}}{(4M)^{1/2s}} \right) \leq \max(\log[C\,(4M)^{k/2s}] - \frac{k}{s} \log[\varepsilon \delta_n^{-s}], 0),$$

as was to be shown.

Finally, using $n^{-1} \sum_{i=1}^n F_n(z_i)^2 = O(\mathbb{E}[F_n(z)^2])$ and proceeding as in the proof of (14) it can be shown that

$$\sup_{\theta \in \Theta(\delta_n), \gamma \in \Gamma_n} \| \hat{G}_n^*(\theta, \gamma) - \hat{G}_n(\theta, \gamma) - \hat{G}_n^*(\theta_0, \gamma) + \hat{G}_n(\theta_0, \gamma) \| = o_p(n^{-1/2 - \alpha s}),$$

a result which is equivalent to (15) when (14) holds.

8.9. Proof of Lemma 9. The proof is analogous to that of Lemma 3. Let $\dot{G}_n^* = \dot{G}(\theta_0, \hat{\gamma}_n^*)$ and define $\mathcal{J}_n^* = -(\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W$. Using $\|\hat{\gamma}_n^* - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, (ii), and (iv*),

$$(\mathcal{J}_n^* - \mathcal{J})\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) = o_p(n^{-1/2}).$$

It therefore suffices to show that $\hat{\theta}_n^* - \hat{\theta}_n - \mathcal{J}_n^*\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) = o_p(n^{-1/2})$.

To do so, let $L_n^*(\theta) = \dot{G}_n^{*\prime}W[\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) + \dot{G}_n^*(\theta - \hat{\theta}_n)]$. Because

$$L_n^*(\hat{\theta}_n^*) = \dot{G}_n^{*\prime}W\dot{G}_n^*[\hat{\theta}_n^* - \hat{\theta}_n - \mathcal{J}_n^*\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)]$$

and because $\dot{G}_n^{*\prime}W\dot{G}_n^* \to_p \dot{G}_0'W\dot{G}_0 > 0$, it suffices to show that $L_n^*(\hat{\theta}_n^*) = o_p(n^{-1/2})$.

Using $\hat{\theta}_n^* - \theta_0 = o_p(1)$, $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$, $\|\hat{\gamma}_n^* - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, (ii), and (iii*) (with $\alpha = 0$),

$$\|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) + \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)\| = o_p(n^{-1/2})$$

and

$$\|\dot{G}_n^*(\hat{\theta}_n^* - \hat{\theta}_n) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) + G(\hat{\theta}_n, \hat{\gamma}_n^*)\| = \|\hat{\theta}_n^* - \theta_0\|^2 O_p(1) + o_p(n^{-1/2}).$$

As a consequence, by the triangle inequality and using $\|\dot{G}_n^{*\prime}W\| = O_p(1)$,

$$\begin{aligned}
\|L_n^*(\hat{\theta}_n^*)\| &\leq \|\dot{G}_n^{*\prime}W\|\|\dot{G}_n^*(\hat{\theta}_n^* - \hat{\theta}_n) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) + G(\hat{\theta}_n, \hat{\gamma}_n^*)\| \\
&\quad + \|\dot{G}_n^{*\prime}W\|\|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) + \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)\| \\
&\quad + \|\dot{G}_n^{*\prime}W\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)\| \\
&= \|\hat{\theta}_n^* - \theta_0\|^2 O_p(1) + \|\dot{G}_n^{*\prime}W\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)\| + o_p(n^{-1/2}),
\end{aligned}$$

so it suffices to show that $\hat{\theta}_n^* - \theta_0 = o_p(n^{-1/3})$ and that $\dot{G}_n^{*\prime} W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_p(n^{-1/2})$.

Proof of $\hat{\theta}_n^* - \theta_0 = o_p(n^{-1/3})$. Condition (ii) implies the existence of a positive constant C for which $\|\theta - \theta_0\| \leq C^{-1} \|G(\theta, \gamma_0)\|$ near θ_0. Because $\hat{\theta}_n^* - \theta_0 = o_p(1)$, it therefore suffices to show that $\|G(\hat{\theta}_n^*, \gamma_0)\| = \|\hat{\theta}_n^* - \theta_0\| o_p(1) + o_p(n^{-1/3})$.

Using (i*), (iv*), and $\hat{\theta}_n \in \Theta$ (with probability approaching one), we have (with probability approaching one)

$$\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \leq \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)' W \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) + o_p(n^{-1}) = o_p(n^{-2/3}),$$

implying in particular that $\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_p(n^{-1/3})$. Also, using $\hat{\theta}_n^* - \theta_0 = o_p(1)$, $\hat{\theta}_n - \theta_0 = o_p(n^{-1/3})$, $\|\hat{\gamma}_n^* - \gamma_0\|_\Gamma = o_p(n^{-1/6})$, (ii), and (iii*) (with $\alpha = 0$),

$$\|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) + \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)\| = o_p(n^{-1/2})$$

and (with probability approaching one)

$$\|G(\hat{\theta}_n^*, \gamma_0) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) + G(\hat{\theta}_n, \hat{\gamma}_n^*)\| \leq \|\hat{\theta}_n^* - \theta_0\| o_p(1) + o_p(n^{-1/3}).$$

Using these rates, the triangle inequality, and (iv*),

$$\begin{aligned} \|G(\hat{\theta}_n^*, \gamma_0)\| &\leq \|G(\hat{\theta}_n^*, \gamma_0) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) + G(\hat{\theta}_n, \hat{\gamma}_n^*)\| \\ &\quad + \|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) + \hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)\| \\ &\quad + \|\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)\| + \|\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)\| \\ &= \|\hat{\theta}_n^* - \theta_0\| o_p(1) + o_p(n^{-1/3}). \end{aligned}$$

Proof of $\dot{G}_n^{*\prime} W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_p(n^{-1/2})$. Because $\dot{G}_n^{*\prime} W \dot{G}_n^* \to_p \dot{G}_0' W \dot{G}_0 > 0$, it suffices to show that

$$\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n)' W \dot{G}_n^* (\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_p(n^{-1}).$$

To do so, let $\tilde{\theta}_n^* = \hat{\theta}_n^* + \mathcal{J}_n^* \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)$, which satisfies $\tilde{\theta}_n^* - \theta_0 = o_p(n^{-1/3})$ because $\hat{\theta}_n^* - \theta_0 = o_p(n^{-1/3})$ and $\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_p(n^{-1/3})$.

By (ii) and (iii*) (with $\alpha = 1/3$), we therefore have

$$R_n^* = \hat{G}_n^*(\tilde{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \dot{G}_n^*(\tilde{\theta}_n^* - \hat{\theta}_n^*) = o_p(n^{-2/3}).$$

As a consequence, using (i) and (v), with probability approaching one

$$\begin{aligned}
\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) &\leq \hat{G}_n^*(\tilde{\theta}_n^*, \hat{\gamma}_n^*)' W \hat{G}_n^*(\tilde{\theta}_n^*, \hat{\gamma}_n^*) + o_p(n^{-1}) \\
&= \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \\
&\quad - \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \dot{G}_n^* (\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \\
&\quad + 2 R_n^{*\prime} [W - W \dot{G}_n^* (\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W] \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \\
&\quad + R_n^{*\prime} W R_n^* + o_p(n^{-1}),
\end{aligned}$$

which rearranges as

$$\begin{aligned}
& \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)' W \dot{G}_n^* (\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \\
\leq\ & 2 R_n^{*\prime} [W - W \dot{G}_n^* (\dot{G}_n^{*\prime} W \dot{G}_n^*)^{-1} \dot{G}_n^{*\prime} W] \hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) + R_n^{*\prime} W R_n^* + o_p(n^{-1}) \\
=\ & o_p(n^{-1}).
\end{aligned}$$

8.10. Proof of Lemma 10. For $i = 1, \ldots, n$, we have

$$\hat{\Psi}_n(X_i) = (1 - n^{-1})\hat{\Psi}_{n,i}(X_i) + \frac{\kappa(0)}{nh_n^d} W_{in}$$

and therefore

$$\max_{1 \leq i \leq n} |\hat{\Psi}_n(X_i) - \Psi_n(X_i)| \leq \max_{1 \leq i \leq n} |\hat{\Psi}_{n,i}(X_i) - \Psi_n(X_i)| + R_n,$$

where

$$R_n = \frac{1}{n} \sup_{x \in \mathbb{R}^d} |\Psi_n(x)| + \frac{C_\kappa}{nh_n^d} \max_{1 \leq i \leq n} |W_{in}| = \frac{C_\kappa}{nh_n^d} \max_{1 \leq i \leq n} |W_{in}| + O(\rho_n)$$

because $n^{-1} \sup_{x \in \mathbb{R}^d} |\Psi_n(x)| \leq n^{-1}\mathcal{C}(1) \int_{\mathbb{R}^d} |\kappa(t)| dt = O(n^{-1}) = O(\rho_n)$. By Chebychev's inequality,

$$\mathbb{P}[\max_{1 \leq i \leq n} |W_{in}| > M\tau_n] \leq n\mathbb{P}[|W_n| > M\tau_n] \leq \frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}}$$

for every M and every (S_n, τ_n). Therefore, $\max_{1 \leq i \leq n} |W_{in}| = O_p(\tau_n)$ if the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by choosing S_n appropriately and making M large.

In case (a), setting $(S_n, \tau_n) = (S, n^{1/S})$ we have $\tau_n/(nh_n^d) = O(\rho_n)$ and

$$\frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{\mathcal{C}(S)}{M^S},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily small by making M large.

In case (b), setting $(S_n, \tau_n) = (\log n, \log n)$ we have $\tau_n/(nh_n^d) = O(\rho_n)$ and

$$\frac{n\mathcal{C}(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{n\mathcal{C}(\log n)}{M^{\log n}(\log n)^{\log n}} \leq \frac{n\Gamma(\log n)H^{\log n}}{M^{\log n}(\log n)^{\log n}} = \left(\frac{H}{M}\right)^{\log n} O(1/\sqrt{\log n}),$$

where the second equality uses Stirling's formula and the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by making M large.

In case (c), setting $(S_n, \tau_n) = (\log n, 1)$ we have $\tau_n/(nh_n^d) = O(\rho_n)$ and

$$\frac{nC_W(S_n)}{M^{S_n}\tau_n^{S_n}} = \frac{nC_W(\log n)}{M^{\log n}} \leq n\left(\frac{H}{M}\right)^{\log n},$$

where the $\overline{\lim}_{n\to\infty}$ of the majorant can be made arbitrarily small by making M large.

In all cases, $R_n = O_p(\rho_n)$ because $\tau_n/(nh_n^d) = O(\rho_n)$. The proof of (16) can therefore be completed by showing that (17) holds.

Proof of (17). With (S_n, τ_n) as before, let

$$\hat{\Psi}^{\tau}_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} W^{\tau}_{jn}\kappa_n(x - X_j), \qquad W^{\tau}_{jn} = W_{jn}1[|W_{jn}| \leq C_\tau\tau_n],$$

where C_τ is a constant to be chosen. We have

$$\mathbb{P}[\hat{\Psi}_{n,i}(\cdot) \neq \hat{\Psi}^{\tau}_{n,i}(\cdot) \text{ for some } i] \leq \mathbb{P}[\max\nolimits_{1\leq i\leq n}|W_{in}| > C_\tau\tau_n],$$

whose $\overline{\lim}_{n\to\infty}\mathbb{P}[\hat{\Psi}_{n,i}(\cdot) \neq \hat{\Psi}^{\tau}_{n,i}(\cdot)$ for some $i]$ can be made arbitrarily small by making C_τ large. Also,

$$\max\nolimits_{1\leq i\leq n}\sup\nolimits_{x\in\mathbb{R}^d}|\mathbb{E}[\hat{\Psi}_{n,i}(x) - \hat{\Psi}^{\tau}_{n,i}(x)]| = O(n^{-1/2}) = O(\rho_n)$$

because

$$\begin{aligned} \frac{n}{\tau_n}|\mathbb{E}[\hat{\Psi}_{n,i}(x) - \hat{\Psi}^{\tau}_{n,i}(x)]| &= \frac{n}{\tau_n}|\mathbb{E}[W_n 1(|W_n| > C_\tau \tau_n)\kappa_n(x - X)]| \\ &\leq \frac{n\mathcal{C}(S_n)}{C_\tau^{S_n}\tau_n^{S_n}} C_\tau \int_{\mathbb{R}^d} |\kappa(t)|dt, \end{aligned}$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily small by making C_τ large. To show the desired result it therefore suffices to show that

$$\max\nolimits_{1\leq i\leq n} |\hat{\Psi}^{\tau}_{n,i}(X_i) - \Psi^{\tau}_n(X_i)| = O_p(\rho_n)$$

for every C_τ, where $\Psi^\tau_n(x) = \mathbb{E}\hat{\Psi}^\tau_n(x) = \mathbb{E}\hat{\Psi}^\tau_{n,i}(x)$.

For any M,

$$\begin{aligned} \mathbb{P}\left[\max\nolimits_{1\leq i\leq n} |\hat{\Psi}^{\tau}_{n,i}(X_i) - \Psi^{\tau}_n(X_i)| > M\rho_n\right] &\leq n \max\nolimits_{1\leq i\leq n} \mathbb{P}[|\hat{\Psi}^{\tau}_{n,i}(X_i) - \Psi^{\tau}_n(X_i)| > M\rho_n] \\ &\leq n \max\nolimits_{1\leq i\leq n} \sup\nolimits_{x\in\mathbb{R}^d} \mathbb{P}[|\hat{\Psi}^{\tau}_{n,i}(x) - \Psi^{\tau}_n(x)| > M\rho_n], \end{aligned}$$

where the last inequality uses the fact that X_i is independent of $\hat{\Psi}^\tau_{n,i}$.

Because

$$|W^\tau_{jn}\kappa_n(x - X_j) - \Psi^\tau_n(x)| = O(h_n^{-d}\tau_n), \qquad \mathbb{V}[W^\tau_{jn}\kappa_n(x - X_j)] = O(h_n^{-d}),$$

it follows from Bernstein's inequality that

$$n \max_{1 \leq i \leq n} \sup_{x \in \mathbb{R}^d} \mathbb{P}[|\hat{\Psi}^{\tau}_{n,i}(x) - \Psi^{\tau}_n(x)| > M\rho_n] \leq 2n \exp\left[-\frac{M^2 n \rho_n^2 h_n^d}{O(1 + M\rho_n\tau_n)}\right].$$

To complete the proof of (17) it therefore suffices to show that

$$\overline{\lim}_{n\to\infty} \frac{1}{\log n} \frac{M^2 n \rho_n^2 h_n^d}{1 + M\rho_n\tau_n}$$

can be made arbitrarily large by making M large.

In case (a), the desired result follows from the proof of Cattaneo et al. (2013, Lemma B-1).

In case (b),

$$\frac{1}{\log n} \frac{M^2 n \rho_n^2 h_n^d}{1 + M\rho_n\tau_n} = \frac{M^2}{1 + MC_\tau \rho_n \log n},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily large (by making M large) if $\rho_n \log n = \sqrt{(\log n)^3/(nh_n^d)}$ is bounded.

In case (c),

$$\frac{1}{\log n} \frac{M^2 n \rho_n^2 h_n^d}{1 + M\rho_n\tau_n} = \frac{M^2}{1 + MC_\tau \rho_n},$$

whose $\overline{\lim}_{n\to\infty}$ can be made arbitrarily large (by making M large) if ρ_n is bounded.

Proof of (18). For any M,

$$\mathbb{P}[\max_{1 \leq i \leq n} |\hat{\Psi}^*_n(X_i) - \hat{\Psi}_n(X_i)| > M\rho_n] = \mathbb{E}\mathbb{P}^*[\max_{1 \leq i \leq n} |\hat{\Psi}^*_n(X_i) - \hat{\Psi}_n(X_i)| > M\rho_n]$$

and

$$\mathbb{P}^*[\max_{1\le i\le n}|\hat{\Psi}_n^*(X_i)-\hat{\Psi}_n(X_i)|>M\rho_n]\le n\sup_{x\in\mathbb{R}^d}\mathbb{P}^*[|\hat{\Psi}_n^*(x)-\hat{\Psi}_n(x)|>M\rho_n].$$

Because

$$|W_{jn}^*\kappa_n(x-X_j^*)-\hat{\Psi}_n(x)|=O_p(h_n^{-d}\tau_n),\qquad \mathbb{V}^*[W_{jn}^*\kappa_n(x-X_j^*)]=O_p(h_n^{-d}),$$

it follows from Bernstein's inequality that

$$\mathbb{P}^*[|\hat{\Psi}_n^*(x)-\hat{\Psi}_n(x)|>M\rho_n]\le 2\exp\left[-\frac{M^2n\rho_n^2h_n^d}{O_p(1+M\rho_n\tau_n)}\right].$$

Validity of (18) follows from this bound and the fact that

$$\overline{\lim}_{n\to\infty}\frac{1}{\log n}\frac{M^2n\rho_n^2h_n^d}{1+M\rho_n\tau_n}$$

can be made arbitrarily large by making M large.

Proof of (19). Because

$$\hat{\Psi}_{n,i}^*(x)=(1-n^{-1})^{-1}\hat{\Psi}_n^*(x)-(n-1)^{-1}W_{in}^*\kappa_n(x-X_i^*),$$

we have the bound

$$(1-n^{-1})\max_{1\le i,j\le n}|\hat{\Psi}_{n,i}^*(X_j)-\hat{\Psi}_n(X_j)|\le\max_{1\le j\le n}|\hat{\Psi}_n^*(X_j)-\hat{\Psi}_n(X_j)|+R_n^*,$$

where

$$
\begin{aligned}
R_n^* &= \frac{1}{n}\max_{1\le i\le n}|\hat{\Psi}_n(X_i)| + \frac{C_\kappa}{nh_n^d}\max_{1\le i\le n}|W_{in}| \\
&\le \frac{1}{n}\max_{1\le i\le n}|\hat{\Psi}_n(X_i) - \Psi_n(X_i)| + \frac{1}{n}\sup_{x\in\mathbb{R}^d}|\Psi_n(x)| + \frac{C_\kappa}{nh_n^d}\max_{1\le i\le n}|W_{in}| = O_p(\rho_n).
\end{aligned}
$$

In particular, (19) holds because (18) holds.

9. Appendix III: Details for the Examples

The purpose of this section is to give explicit regularity conditions and to provide details on the derivations of the results for Examples 1-3.

9.1. Example 1. To obtain primitive bandwidth conditions for the conditions of Theorems 2 and 7, suppose that for some $P > d/2$,

- f_0 is P times differentiable, and f_0 and its first P derivatives are bounded and continuous.
- K is even and bounded with $\int_{\mathbb{R}^d} |K(u)| (1 + \|u\|^P) du < \infty$ and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_d = 0, \\ 0, & \text{if } (l_1, \ldots, l_d)' \in \mathbb{Z}_+^d \text{ and } l_1 + \cdots + l_d < P \end{cases}.$$

Conditions (AL) and (AL*) hold with $\mathcal{J} = I_k$ and without any $o_p(1)$ terms.

Condition (AS). Because $g_n(x, \cdot)$ is linear, Condition (AS) holds with $\bar{g}_n = g_n$ if $\mathbb{V}(g_{n,\gamma}(z_1)[\kappa_{n,2}]) = o(n)$. A sufficient condition for this occur is that $nh_n^d \to \infty$, because then

$$\mathbb{V}(g_{n,\gamma}(z_1)[\kappa_{n,2}]) = (1 - n^{-1})^2 \mathbb{V}[\mathcal{K}_n(x_1 - x_2) - f_n(x_1)] = O(h_n^{-d}) = o(n).$$

Condition (SE). If $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n [\hat{f}_n(x_i) - f_n(x_i) - f_0(x_i) + \theta_0] \\
= \ &\frac{1}{\sqrt{n}} \sum_{i=1}^n [n^{-1}\mathcal{K}_n(0) + (1 - n^{-1})\hat{f}_{n,i}(x_i) - f_n(x_i) - f_0(x_i) + \theta_0] \\
= \ &\frac{K(0)}{\sqrt{nh_n^{2d}}} + \frac{1}{\sqrt{n}} \sum_{i=1}^n [(1 - n^{-1})(2f_n(x_i) - \theta_n) - f_n(x_i) - f_0(x_i) + \theta_0] + o_p(1) \\
= \ &\frac{K(0)}{\sqrt{nh_n^{2d}}} + \frac{1}{\sqrt{n}} \sum_{i=1}^n [f_n(x_i) - f_0(x_i)] - \sqrt{n}(\theta_n - \theta_0) + o_p(1) = \frac{K(0)}{\sqrt{nh_n^{2d}}} + o_p(1),
\end{aligned}$$

where the last equality uses $\theta_n - \theta_0 = O(h_n^P)$ and $\mathbb{E}(|f_n(x) - f_0(x)|^2) = o(1)$, and the second equality uses Condition (AS). As a consequence, Condition (SE) requires $nh_n^{2d} \to \infty$.

Condition (AN). We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(x_i, f_n) + \bar{G}_n(\hat{f}_{n,i}) - \bar{G}_n(f_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n(x_i) + B_n],$$

where

$$\psi_n(x) = 2[f_n^+(x) - \theta_n^+], \qquad \theta_n^+ = \int_{\mathbb{R}^d} f_n^+(x) f_0(x) dx,$$

and $B_n = \theta_n^+ - \theta_0$. If $h_n \to 0$, then $\psi_n(x) \to \psi(x)$ for every x and it follows from the dominated convergence theorem that (9) is satisfied. Also, (10) is satisfied because $B_n = K(0)/(nh_n^d) + O(h_n^P + n^{-1})$ is nonrandom. Condition (AN) is therefore satisfied (with $\Sigma = 4\mathbb{V}[f_0(x)]$) if $h_n \to 0$ and, in fact, we can take $\beta_n = K(0)/(nh_n^d)$ when $nh_n^{2P} \to 0$.

Condition (AS).* Because $g_n^*(x, \cdot)$ is linear, Condition (AS*) holds with $\bar{g}_n^* = g_n^*$ if $\mathbb{V}^*(\bar{g}_{n,\gamma}^*(z_1^*)[\kappa_{n,2}^*]) = o_p(n)$. A sufficient condition for this to occur is that $nh_n^d \to \infty$, because then

$$\mathbb{E}\mathbb{V}^*(g_{n,\gamma}^*(z_1^*)[\kappa_{n,2}^*]) = (1 - n^{-1})^2 \mathbb{E}\mathbb{V}^*[\mathcal{K}_n(x_1^* - x_2^*) - \hat{f}_n(x_1^*)] = O(h_n^{-d}) = o(n).$$

Condition (AN).* Finally, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(x_i^*, \hat{f}_n) + \bar{G}_n^*(\hat{f}_{n,i}^*) - \bar{G}_n^*(\hat{f}_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n^*(x_i^*) + B_n^*],$$

where, defining $\hat{f}_n^+(x) = n^{-1}\mathcal{K}_n(0) + (1 - n^{-1})\hat{f}_n(x)$,

$$\psi_n^*(x) = 2[\hat{f}_n^+(x) - \hat{\theta}_n^+], \qquad \hat{\theta}_n^+ = \frac{1}{n}\sum_{i=1}^n \hat{f}_n^+(x_i), \qquad B_n^* = n^{-1}\mathcal{K}_n(0) - n^{-1}\hat{\theta}_n.$$

Suppose $h_n \to 0$ and $nh_n^d \to \infty$. Then $B_n^* = n^{-1}\mathcal{K}_n(0) - n^{-1}\hat{\theta}_n = \beta_n + o_p(n^{-1/2})$ because $\hat{\theta}_n = O_p(1)$. Because $\hat{\theta}_n - \theta_n \to_p 0$, $n^{-1}\sum_{i=1}^n |\psi_n^*(x_i) - \psi_n(x_i)|^2 \to_p 0$ also holds provided

$$\frac{1}{n}\sum_{i=1}^n |\hat{f}_n(x_i) - f_n(x_i)|^2 \to_p 0.$$

A sufficient condition for this to occur is that $\max_i |\hat{f}_n(x_i) - f_n(x_i)| = o_p(1)$, which in turn will hold if $nh_n^d/\log n \to \infty$. Sufficiency of the slightly weaker condition $nh_n^d \to \infty$ can be demonstrated by using a direct calculation to show that if $nh_n^d \to \infty$, then

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n |\hat{f}_n(x_i) - f_n(x_i)|^2\right) = O\left(\frac{1}{nh_n^d}\right) \to 0.$$

In other words, Condition (AN*) holds if $h_n \to 0$ and if $nh_n^d \to \infty$.

9.2. Example 2. To obtain primitive bandwidth conditions for the conditions of Theorems 2 and 7, suppose that for some $P > d \geq 3$,

- $\mathbb{E}[\,|y|^4] + \sup_x \mathbb{E}[|y|^4|x]f_0(x) < \infty$.
- $\Sigma = \mathbb{V}[\psi(z)]$ is positive definite.
- ω is continuously differentiable, and ω and its first derivative are bounded.

- $\inf_{x:\omega(x)>0} f_0(x) > 0$.
- f_0 is $P+1$ times differentiable, and f_0 and its first $P+1$ derivatives are bounded and continuous.
- γ_r is continuously differentiable, and γ_r and its first derivative are bounded, where $\gamma_r(x) = r(x)f_0(x)$.
- $\lim_{\|x\|\to\infty}[f_0(x) + |\gamma_r(x)|] = 0$, where $\|\cdot\|$ is the Euclidean norm.
- K is even and differentiable, and K and its first derivative are bounded.
- $\int_{\mathbb{R}^d} \|\partial K(u)/\partial u\| du + \int_{\mathbb{R}^d} |K(u)|(1 + \|u\|^P) du < \infty$ and

$$\int_{\mathbb{R}^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_d = 0, \\ 0, & \text{if } (l_1, \ldots, l_d)' \in \mathbb{Z}_+^d \text{ and } l_1 + \cdots + l_d < P \end{cases}.$$

Conditions (AL) and (AL*) hold with $\mathcal{J} = I_k$ and without any $o_p(1)$ terms.

Condition (AS). Suppose $h_n \to 0$. As in the text, let

$$\bar{g}_n(z,f) = g_n(z,f_n) + g_{n,f}(z)[f - f_n] + \frac{1}{2} g_{n,ff}(z)[f - f_n, f - f_n],$$

where, defining $f_n^+(x) = n^{-1}\mathcal{K}_n(0) + (1 - n^{-1}) f_n(x)$,

$$\begin{aligned} g_{n,f}(z)[\kappa] &= -(1-n^{-1}) \frac{y\omega(x)}{f_n^+(x)} \left[\frac{\partial}{\partial x}\kappa(x) - \frac{\partial f_n^+(x)/\partial x}{f_n^+(x)} \kappa(x) \right], \\ g_{n,ff}(z)[\kappa,\lambda] &= (1-n^{-1})^2 \frac{y\omega(x)}{f_n^+(x)^2} \left[\lambda(x)\frac{\partial}{\partial x}\kappa(x) + \kappa(x)\frac{\partial}{\partial x}\lambda(x) - 2\frac{\partial f_n^+(x)/\partial x}{f_n^+(x)} \kappa(x)\lambda(x) \right]. \end{aligned}$$

The assumptions of Lemma 5 are satisfied if $\max(\Delta_n, \dot{\Delta}_n) = o_p(n^{-1/6})$, where

$$\Delta_n = \max_i \left| \hat{f}_{n,i}(x_i) - f_n(x_i) \right|, \qquad \dot{\Delta}_n = \max_i \left| \frac{\partial}{\partial x_i} \hat{f}_{n,i}(x_i) - \frac{\partial}{\partial x_i} f_n(x_i) \right|.$$

More generally, proceeding as in Cattaneo et al. (2013) it can be shown that the first part of Condition (AS) is satisfied if $\Delta_n = o_p(1)$ and if $\Delta_n^2 \max(\Delta_n, \dot{\Delta}_n) = o_p(n^{-1/2})$. Because (by Lemma 10)

$$\Delta_n = O_p(1/\sqrt{nh_n^d/\log n}), \qquad \dot{\Delta}_n = O_p(1/\sqrt{nh_n^{d+2}/\log n}),$$

a sufficient condition for this to occur is that $nh_n^{\frac{3}{2}d+1}/\left(\log n\right)^{3/2} \to \infty$. Moreover,

$$\mathbb{V}(g_{n,f}(z_1)[\kappa_{n,2}]) = O(h_n^{-d-2}),$$
$$\mathbb{V}[\mathbb{E}(g_{n,ff}(z_1)[\kappa_{n,2},\kappa_{n,2}]|z_1)] = O(h_n^{-2d-2}), \qquad \mathbb{V}(g_{n,ff}(z_1)[\kappa_{n,2},\kappa_{n,2}]) = O(h_n^{-3d-2}),$$
$$\mathbb{V}(g_{n,ff}(z_1)[\kappa_{n,2},\kappa_{n,3}]) = O(h_n^{-2d-2}),$$

so the assumptions of Lemma 4 will be satisfied provided $nh_n^{d+2} \to \infty$.

Condition (AN). We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i, f_n) + \bar{G}_n(\hat{f}_{n,i}) - \bar{G}_n(f_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n(z_i) + B_n],$$

where

$$\psi_n(z) = g_n(z, f_n) - \mathbb{E}g_n(z, f_n) + \delta_n(z),$$
$$\delta_n(z) = (1 - n^{-1})\int_{\mathbb{R}^d} \delta_0(s)\frac{f_0(s)}{f_n^+(s)}[\mathcal{K}_n(s-x) - f_n(s)]ds,$$
$$\delta_0(z) = \omega(x)\frac{\partial}{\partial x}r(x) + r(x)\frac{\partial}{\partial x}\omega(x) + r(x)\omega(x)\frac{\partial f_0(x)/\partial x}{f_0(x)},$$

and

$$B_n = \mathbb{E}g_n(z, f_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} G_{n,ff}[\hat{f}_{n,i} - f_n, \hat{f}_{n,i} - f_n].$$

If $h_n \to 0$ and if $nh_n^d \to \infty$, then $\psi_n(z) \to \psi(z) = g_0(z, f_0) + \delta_0(z)$ for every z and it follows from the dominated convergence theorem that (9) is satisfied. Also, (10) is satisfied if $h_n \to 0$ and if $nh_n^{d+2} \to \infty$ because the representation

$$\begin{aligned} G_{n,ff}[\kappa, \lambda] &= (1 - n^{-1})^2 \int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_n^+(x)^2}[\lambda(x)\frac{\partial}{\partial x}\kappa(x) + \kappa(x)\frac{\partial}{\partial x}\lambda(x)]f_0(x)dx \\ &\quad -2(1 - n^{-1})^2 \int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_n^+(x)^2}\frac{\partial f_n^+(x)/\partial x}{f_n^+(x)}\kappa(x)\lambda(x)f_0(x)dx \end{aligned}$$

can be used to show that if $h_n \to 0$, then

$$\mathbb{E}(\|G_{n,ff}[\kappa_{n,1}, \kappa_{n,1}]\|^2) = O(h_n^{-2d-2}), \qquad \mathbb{E}(\|G_{n,ff}[\kappa_{n,1}, \kappa_{n,2}]\|^2) = O(h_n^{-d-2}).$$

Condition (AN) is therefore satisfied (with $\Sigma = \mathbb{E}[\psi(z)\psi(z)']$) if $h_n \to 0$ and if $nh_n^{d+2} \to \infty$. With some additional effort it can be shown that if also $nh_n^{2P} \to 0$, then $\mathbb{E}B_n = \beta_n + o(n^{-1/2})$, where

$$\begin{aligned} \beta_n &= \frac{1}{nh_n^d}K(0)\int_{\mathbb{R}^d} r(x)\omega(x)\frac{\partial f_0(x)/\partial x}{f_0(x)}dx \\ &\quad +\frac{1}{nh_n^{d+1}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_0(x)}K(r)\frac{\partial}{\partial r}K(r)f_0(x - rh_n)dxdr \\ &\quad -\frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{r(x)\omega(x)}{f_0(x)}\frac{\partial f_0(x)/\partial x}{f_0(x)}K(r)^2 f_0(x - rh_n)dxdr. \end{aligned}$$

Condition (AS).* A quadratic approximation to $g_n^*(z,f)$ is given by

$$\bar{g}_n^*(z,f) = g_n^*(z,\hat{f}_n) + g_{n,f}^*(z)[f-\hat{f}_n] + \frac{1}{2} g_{n,ff}^*(z)[f-\hat{f}_n, f-\hat{f}_n],$$

where

$$\begin{aligned} g_{n,f}^*(z)[\kappa] &= -(1-n^{-1})\frac{y\omega(x)}{\hat{f}_n^+(x)}\left[\frac{\partial}{\partial x}\kappa(x) - \frac{\partial \hat{f}_n^+(x)/\partial x}{\hat{f}_n^+(x)}\kappa(x)\right], \\ g_{n,ff}^*(z)[\kappa,\lambda] &= (1-n^{-1})^2\frac{y\omega(x)}{\hat{f}_n^+(x)^2}\left[\lambda(x)\frac{\partial}{\partial x}\kappa(x) + \kappa(x)\frac{\partial}{\partial x}\lambda(x) - 2\frac{\partial \hat{f}_n^+(x)/\partial x}{\hat{f}_n^+(x)}\kappa(x)\lambda(x)\right]. \end{aligned}$$

Suppose $h_n \to 0$ and $\Delta_n = o_p(1)$. Then the first part of Condition (AS*) is satisfied if $\Delta_n^* = o_p(1)$ and if $\Delta_n^{*2}\max(\Delta_n^*, \dot{\Delta}_n^*) = o_p(n^{-1/2})$, where

$$\Delta_n^* = \max_i |\hat{f}_{n,i}^*(x_i^*) - \hat{f}_n(x_i^*)|, \qquad \dot{\Delta}_n^* = \max_i |\frac{\partial}{\partial x}\hat{f}_{n,i}^*(x_i^*) - \frac{\partial}{\partial x}\hat{f}_n(x_i^*)|.$$

Because (by Lemma 10)

$$\Delta_n^* = O_p(1/\sqrt{nh_n^d/\log n}), \qquad \dot{\Delta}_n^* = O_p(1/\sqrt{nh_n^{d+2}/\log n}),$$

a sufficient condition for this to occur is that $nh_n^{\frac{3}{2}d+1}/(\log n)^{3/2} \to \infty$. Moreover, if $nh_n^d \to \infty$, then

$$\mathbb{V}^*(g_{n,f}^*(z_1^*)[\kappa_{n,2}^*]) = O_p(h_n^{-d-2}),$$

$$\mathbb{V}^*[\mathbb{E}^*(g_{n,ff}^*(z_1^*)[\kappa_{n,2}^*,\kappa_{n,2}^*]|z_1^*)] = O_p(h_n^{-2d-2}), \qquad \mathbb{V}^*(g_{n,ff}^*(z_1^*)[\kappa_{n,2}^*,\kappa_{n,2}^*]) = O_p(h_n^{-3d-2}),,$$

$$\mathbb{V}^*(g^*_{n,ff}(z^*_1)[\kappa^*_{n,2},\kappa^*_{n,3}]) = O_p(h_n^{-2d-2}),$$

so the second part of Condition (AS*) will be satisfied provided $nh_n^{d+2} \to \infty$.

Condition (AN).* Finally, we have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g^*_n(z^*_i,\hat{f}_n) + \bar{G}^*_n(\hat{f}^*_{n,i}) - \bar{G}^*_n(\hat{f}_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi^*_n(z^*_i) + B^*_n],$$

where

$$\psi^*_n(z) = g^*_n(z,\hat{f}_n) - \frac{1}{n}\sum_{i=1}^{n} g^*_n(z_i,\hat{f}_n) + \delta^*_n(z),$$

$$\begin{aligned}
\delta^*_n(z) &= -(1-n^{-1})\frac{1}{n}\sum_{i=1}^{n}\frac{y_i\omega(x_i)}{\hat{f}^+_n(x_i)}\left[\frac{\partial}{\partial x_i}\dot{\mathcal{K}}_n(x_i-x) - \frac{\partial}{\partial x_i}\hat{f}_n(x_i)\right] \\
&\quad +(1-n^{-1})\frac{1}{n}\sum_{i=1}^{n}\frac{y_i\omega(x_i)}{\hat{f}^+_n(x_i)}\frac{\partial\hat{f}^+_n(x_i)/\partial x_i}{\hat{f}^+_n(x_i)}[\mathcal{K}_n(x_i-x) - \hat{f}_n(x_i)].
\end{aligned}$$

$$B^*_n = \frac{1}{n}\sum_{i=1}^{n} g^*_n(z_i,\hat{f}_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} G^*_{n,ff}[\hat{f}^*_{n,i}-\hat{f}_n,\hat{f}^*_{n,i}-\hat{f}_n].$$

Suppose $h_n \to 0$ and $nh_n^{d+2}/(\log n) \to \infty$. Using Lemma 10 and the fact that $\hat{\theta}_n \to_p \theta_0$ it can be shown that $n^{-1}\sum_{i=1}^{n}\|\psi^*_n(z_i) - \psi_n(z_i)\|^2 \to_p 0$. Also, the bootstrap analog of (10) is satisfied because the representation

$$\begin{aligned}
G^*_{n,ff}[\kappa,\lambda] &= \left(1-\frac{1}{n}\right)^2\frac{1}{n}\sum_{i=1}^{n}\frac{y_i\omega(x_i)}{\hat{f}^+_n(x_i)^2}\left[\dot{\kappa}(x_i)\lambda(x_i) + \kappa(x_i)\dot{\lambda}(x_i)\right] \\
&\quad -2\left(1-\frac{1}{n}\right)^2\frac{1}{n}\sum_{i=1}^{n}\frac{y_i\omega(x_i)}{\hat{f}^+_n(x_i)^2}\left[\frac{\partial\hat{f}^+_n(x_i)/\partial x_i}{\hat{f}^+_n(x_i)}\kappa(x_i)\lambda(x_i)\right]
\end{aligned}$$

can be used to show that

$$\mathbb{V}^*(\|G^*_{n,ff}[\kappa^*_{n,1},\kappa^*_{n,1}]\|^2) = O_p(h_n^{-2d-2}), \qquad \mathbb{V}^*(\|G^*_{n,ff}[\kappa^*_{n,1},\kappa^*_{n,2}]\|^2) = O_p(h_n^{-d-2}).$$

Finally, it can be shown with some effort that $\mathbb{E}^*(B^*_n) = \beta_n + o_p\left(n^{-1/2}\right)$ if $nh_n^{\frac{3}{2}d+1}/\left(\log n\right)^{3/2} \to \infty$. In other words, Condition (AN*) holds if $h_n \to 0$ and if $nh_n^{\frac{3}{2}d+1}/\left(\log n\right)^{3/2} \to \infty$.

9.3. Example 3. To obtain primitive bandwidth conditions for the conditions of Theorems 2 and 7, suppose that for some $P > 3d/4$,

- $F_{y|x}(\cdot|x)$, the conditional cdf of y given x, has three bounded (uniformly in x) derivatives.
- K is even and bounded with $\int_{\mathbb{R}^d}|K(u)|(1+\|u\|^P)du < \infty$ and

$$\int_{\mathbb{R}^d} u_1^{l_1}\cdots u_d^{l_d}K(u)du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_d = 0, \\ 0, & \text{if } (l_1,\ldots,l_d)' \in \mathbb{Z}^d_+ \text{ and } l_1+\cdots+l_d < P \end{cases}.$$

Conditions (AL) and (AL*) hold with $\mathcal{J} = I_k$ and without any $o_p(1)$ terms.

Condition (AS). Let $F_{y|x}(\cdot|x)$ denote the conditional cdf of y given x, let $f_{y|x}(\cdot|x)$ and $\dot{f}_{y|x}(\cdot|x)$ denote its first and second derivatives, and define

$$\breve{g}_n(x,f) = \mathbb{E}[g_n(z,f)|x] - (1-\theta_0) = -F_{y|x}[n^{-1}\mathcal{K}_n(0) + (1-n^{-1})f(x)|x].$$

Being a defined through a projection, $\breve{g}_n(x,f)$ is likely to be close to $g_n(z,f)$ in the the appropriate sense and, indeed,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\hat{f}_{n,i})-\breve{g}_n(x_i,\hat{f}_{n,i})-g_n(z_i,f_n)+\breve{g}_n(x_i,f_n)]=o_p(1)$$

if $\Delta_n = o_p(1)$, because then

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i,\hat{f}_{n,i})-\breve{g}_n(x_i,\hat{f}_{n,i})-g_n(z_i,f_n)+\breve{g}_n(x_i,f_n)]\right)^2|\mathcal{X}_n\right]\\ = \;& \frac{1}{n}\mathbb{V}\left(\sum_{i=1}^{n}[g_n(z_i,\hat{f}_{n,i})-g_n(z_i,f_n)]|\mathcal{X}_n\right)\leq \sup\nolimits_{r,s} f_{y|x}(r|s)\Delta_n = o_p(1),\end{aligned}$$

where $\mathcal{X}_n = (x_1,\ldots,x_n)'$. Next, being smooth $\breve{g}_n(x,f)$ admits the quadratic approximation

$$\bar{g}_n(x,f)=\breve{g}_n(x,f_n)+\breve{g}_{n,f}(x)[f-f_n]+\frac{1}{2}\breve{g}_{n,ff}(x)[f-f_n,f-f_n],$$

where

$$\begin{aligned}\breve{g}_{n,f}(x)[\kappa] &= -(1-n^{-1})f_{y|x}[f_n^+(x)|x]\kappa(x),\\ \breve{g}_{n,ff}(x)[\kappa,\lambda] &= -(1-n^{-1})^2\dot{f}_{y|x}[f_n^+(x)|x]\kappa(x)\lambda(x).\end{aligned}$$

It follows from standard bounding arguments (e.g., Lemma 5) that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\breve{g}_n(x_i,\hat{f}_{n,i})-\bar{g}_n(x_i,\hat{f}_{n,i})-\breve{g}_n(x_i,f_n)+\bar{g}_n(x_i,f_n)]=o_p(1)$$

provided $\Delta_n = o_p(n^{-1/6})$. This condition, and therefore the first part of Condition (AS), is satisfied when $nh_n^{\frac{3}{2}d}/(\log n)^{3/2}\to\infty$. Moreover,

$$\mathbb{V}(\breve{g}_{n,f}(x_1)[\kappa_{n,2}]) = O(h_n^{-d}),$$
$$\mathbb{V}[\mathbb{E}(\breve{g}_{n,ff}(x_1)[\kappa_{n,2},\kappa_{n,2}]|z_1)] = O(h_n^{-2d}), \qquad \mathbb{V}(\breve{g}_{n,ff}(x_1)[\kappa_{n,2},\kappa_{n,2}]) = O(h_n^{-3d}),$$
$$\mathbb{V}(\breve{g}_{n,ff}(x_1)[\kappa_{n,2},\kappa_{n,3}]) = O(h_n^{-2d}),$$

so the assumptions of Lemma 4 will be satisfied provided $nh_n^d \to \infty$.

Condition (AN). We have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n(z_i, f_n) + \bar{G}_n(\hat{f}_{n,i}) - \bar{G}_n(f_n)] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n(z_i) + B_n],$$

where

$$\psi_n(z) = g_n(z, f_n) - \mathbb{E}g_n(z, f_n) + \delta_n(x),$$
$$\delta_n(x) = -(1-n^{-1})\int_{\mathbb{R}^d} f_{y|x}[f_n^+(r)|r][\mathcal{K}_n(r-x) - f_n(r)]f_0(r)dr,$$

and

$$B_n = \mathbb{E}g_n(z, f_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\breve{G}_{n,ff}[\hat{f}_{n,i} - f_n, \hat{f}_{n,i} - f_n].$$

If $h_n \to 0$ and if $nh_n^d \to \infty$, then

$$\psi_n(z) \to \psi(z) = g_0(z, f_0) + \delta_0(x), \qquad \delta_0(x) = -f_{y|x}[f_0(x)|x]f_0(x) + \int_{\mathbb{R}^d} f_{y|x}[f_0(x)|x]f_0(x)^2 dx,$$

for every z and it follows from the dominated convergence theorem that (9) is satisfied. Also, (10) is satisfied if $h_n \to 0$ and if $nh_n^d \to \infty$ because the representation

$$\breve{G}_{n,ff}[\kappa,\lambda] = -(1-n^{-1})^2 \int_{\mathbb{R}^d} \dot{f}_{y|x}[f_n^+(x)|x]\kappa(x)\lambda(x)f_0(x)dx$$

can be used to show that if $h_n \to 0$, then

$$\mathbb{E}(\|\breve{G}_{n,ff}[\kappa_{n,1},\kappa_{n,1}]\|^2) = O(h_n^{-2d}), \qquad \mathbb{E}(\|\breve{G}_{n,ff}[\kappa_{n,1},\kappa_{n,2}]\|^2) = O(h_n^{-d}).$$

Condition (AN) is therefore satisfied (with $\Sigma = \mathbb{E}[\psi(z)\psi(z)']$) if $h_n \to 0$ and if $nh_n^d \to \infty$. With some additional effort it can be shown that if also $nh_n^{2P} \to 0$, then $\mathbb{E}B_n = \beta_n + o(n^{-1/2})$, where

$$\begin{aligned}\beta_n &= -\frac{1}{nh_n^d} K(0) \int_{\mathbb{R}^d} f_{y|x}[f_0(x)|x] f_0(x)\, dx \\ &\quad -\frac{1}{nh_n^d}\frac{1}{2}\int_{\mathbb{R}^d} \dot{f}_{y|x}[f_0(x)|x] K(r)^2 f_0(x) f_0(x - rh_n)\, dxdr.\end{aligned}$$

Condition (AS).* Let $\breve{g}_n^*(x,f) = \breve{g}_n(x,f)$ and define

$$\bar{g}_n^*(x,f) = \breve{g}_n^*(x,\hat{f}_n) + \breve{g}_{n,f}^*(x)[f-\hat{f}_n] + \frac{1}{2}\breve{g}_{n,ff}^*(x)[f-\hat{f}_n, f-\hat{f}_n],$$

where

$$\begin{aligned}\breve{g}_{n,f}^*(x)[\kappa] &= -(1-n^{-1}) f_{y|x}[\hat{f}_n^+(x)|x]\kappa(x), \\ \breve{g}_{n,ff}^*(x)[\kappa,\lambda] &= -(1-n^{-1})^2 \dot{f}_{y|x}[\hat{f}_n^+(x)|x]\kappa(x)\lambda(x).\end{aligned}$$

Defining $N_i = \sum_{j=1}^n 1\left(x_j^* = x_i\right)$ and using the fact (about the multinomial distribution) that $n^{-1}\sum_{i=1}^n N_i^2 = O_p(1)$, it can be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{f}_{n,i}^*)-\breve{g}_n^*(x_i^*,\hat{f}_{n,i}^*)-g_n^*(z_i^*,\hat{f}_n)+\breve{g}_n^*(x_i^*,\hat{f}_n)]=o_p(1)$$

if $\Delta_n^*=o_p\left(1\right)$, because then

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{f}_{n,i}^*)-\breve{g}_n^*(x_i^*,\hat{f}_{n,i}^*)-g_n^*(z_i^*,\hat{f}_n)+\breve{g}_n^*(x_i^*,\hat{f}_n)]\right)^2|\mathcal{X}_n,\mathcal{X}_n^*\right]\\ =\ &\frac{1}{n}\mathbb{V}\left(\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{f}_{n,i}^*)-g_n^*(z_i^*,\hat{f}_n)]|\mathcal{X}_n,\mathcal{X}_n^*\right)\leq \sup\nolimits_{r,s}f_{y|x}(r|s)\left(\frac{1}{n}\sum_{i=1}^{n}N_i^2\right)\Delta_n^*=o_p(1),\end{aligned}$$

where $\mathcal{X}_n^*=(x_1^*,\ldots,x_n^*)'$. Also, it follows from standard bounding arguments that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\breve{g}_n^*(x_i^*,\hat{f}_{n,i}^*)-\bar{g}_n^*(x_i^*,\hat{f}_{n,i}^*)-\breve{g}_n^*(x_i^*,\hat{f}_n)+\bar{g}_n^*(x_i^*,\hat{f}_n)]=o_p(1)$$

provided $\Delta_n^*=o_p(n^{-1/6})$. This condition, and therefore the first part of Condition (AS*), is satisfied when $nh_n^{\frac{3}{2}d}/\left(\log n\right)^{3/2}\to\infty$. Moreover,

$$\mathbb{V}^*(\breve{g}_{n,f}^*(x_1^*)[\kappa_{n,2}^*])=O_p(h_n^{-d}),$$

$$\mathbb{V}^*[\mathbb{E}^*(\breve{g}_{n,ff}^*(x_1^*)[\kappa_{n,2}^*,\kappa_{n,2}^*]|x_1^*)]=O_p(h_n^{-2d}),\qquad \mathbb{V}^*(\breve{g}_{n,ff}^*(x_1^*)[\kappa_{n,2}^*,\kappa_{n,2}^*])=O_p(h_n^{-3d}),$$

$$\mathbb{V}^*(\breve{g}_{n,ff}^*(x_1^*)[\kappa_{n,2}^*,\kappa_{n,3}^*])=O_p(h_n^{-2d}),$$

so the second part of Condition (AS*) will be satisfied provided $nh_n^d\to\infty$.

Condition (AN).* Finally, we have:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g_n^*(z_i^*,\hat{f}_n)+\bar{G}_n^*(\hat{f}_{n,i}^*)-\bar{G}_n^*(\hat{f}_n)]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n^*(z_i^*)+B_n^*],$$

where

$$\psi_n^*(z) = g_n^*(z, \hat{f}_n) - \frac{1}{n}\sum_{i=1}^n g_n^*(z_i, \hat{f}_n) + \delta_n^*(z),$$

$$\delta_n^*(z) = -(1-n^{-1})\frac{1}{n}\sum_{i=1}^n f_{y|x}[\hat{f}_n^+(x_i)|x_i][\mathcal{K}_n(x_i - x) - \hat{f}_n(x_i)],$$

$$B_n^* = \frac{1}{n}\sum_{i=1}^n g_n^*(z_i, \hat{f}_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^n \breve{G}_{n,ff}^*[\hat{f}_{n,i}^* - \hat{f}_n, \hat{f}_{n,i}^* - \hat{f}_n].$$

Suppose $h_n \to 0$ and $nh_n^{\frac{3}{2}d}/\left(\log n\right)^{3/2} \to \infty$. Using Lemma 10 and the fact that $\hat{\theta}_n \to_p \theta_0$ it can be shown that $n^{-1}\sum_{i=1}^n \|\psi_n^*(z_i) - \psi_n(z_i)\|^2 \to_p 0$. Also, the bootstrap analog of (10) is satisfied because the representation

$$\breve{G}_{n,ff}^*[\kappa, \lambda] = -(1-n^{-1})^2\frac{1}{n}\sum_{i=1}^n \dot{f}_{y|x}[\hat{f}_n^+(x_i)|x_i]\kappa(x_i)\lambda(x_i)$$

can be used to show that

$$\mathbb{V}^*(\|\breve{G}_{n,ff}^*[\kappa_{n,1}^*, \kappa_{n,1}^*]\|^2) = O_p(h_n^{-2d}), \qquad \mathbb{V}^*(\|\breve{G}_{n,ff}^*[\kappa_{n,1}^*, \kappa_{n,2}^*]\|^2) = O_p(h_n^{-d}).$$

Finally, it can be shown with some effort that $\mathbb{E}^*(B_n^*) = \beta_n + o_p\left(n^{-1/2}\right)$ if $nh_n^{\frac{3}{2}d}/\left(\log n\right)^{3/2} \to \infty$. In other words, Condition (AN*) holds if $h_n \to 0$ and if $nh_n^{\frac{3}{2}d}/\left(\log n\right)^{3/2} \to \infty$.