Econometrica, Vol. 86, No. 3 (May, 2018), 955–995

KERNEL-BASED SEMIPARAMETRIC ESTIMATORS: SMALL BANDWIDTH ASYMPTOTICS AND BOOTSTRAP CONSISTENCY

MATIAS D. CATTANEO
Department of Economics and Department of Statistics, University of Michigan

MICHAEL JANSSON
Department of Economics, UC Berkeley and CREATES

This paper develops asymptotic approximations for kernel-based semiparametric estimators under assumptions accommodating slower-than-usual rates of convergence of their nonparametric ingredients. Our first main result is a distributional approximation for semiparametric estimators that differs from existing approximations by accounting for a bias. This bias is nonnegligible in general, and therefore poses a challenge for inference. Our second main result shows that some (but not all) nonparametric bootstrap distributional approximations provide an automatic method of correcting for the bias. Our general theory is illustrated by means of examples and its main finite sample implications are corroborated in a simulation study.

KEYWORDS: Semiparametrics, small bandwidth asymptotics, bootstrapping, robust inference.

1. INTRODUCTION

THE IMPORTANCE OF SEMIPARAMETRIC ESTIMATORS is widely recognized, yet the consensus opinion seems to be that existing large sample results suffer from the serious shortcoming that the finite sample distributions of these 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 that are important in samples of realistic size. Motivated by this observation, and with the ultimate goal of developing more "robust" inference procedures based on semiparametric estimators, this paper obtains two main results. (We employ a certain well-defined sense of "robustness" discussed precisely below.)

First, we revisit the large sample properties of kernel-based semiparametric estimators and obtain novel distributional approximations for members of this large class. By design, these approximations capture certain features of their nonparametric ingredient that are ignored by conventional approximations. Moreover, as a consequence of their method

Matias D. Cattaneo: cattaneo@umich.edu
Michael Jansson: mjansson@econ.berkeley.edu

A previous version of this paper was circulated under the title "Bootstrapping Kernel-Based Semiparametric Estimators." We thank the handling coeditor, four referees, Stephane Bonhomme, Lutz Kilian, Xinwei Ma, Whitney Newey, Jamie Robins, Adam Rosen, Andres Santos, Azeem Shaikh, Xiaoxia Shi, and seminar participants at Cambridge University, Georgetown University, London School of Economics, Oxford University, Rice University, University of Chicago, University College London, University of Michigan, the 2013 Latin American Meetings of the Econometric Society, and 2014 CEME/NSF Conference on Inference in Nonstandard Problems for comments. The first author gratefully acknowledges financial support from the National Science Foundation (SES 1122994 and SES 1459931). The second author gratefully acknowledges financial support from the National Science Foundation (SES 1124174 and SES 1459967) and the research support of *CREATES* (funded by the Danish National Research Foundation under Grant no. DNRF78).

 https://doi.org/10.3982/ECTA12701

of construction, our approximations are demonstrably more robust than conventional ones in the sense that we allow for (but do not require) nonparametric ingredients whose precision is low enough (in an order of magnitude sense) to render conventional distributional approximations invalid. Accordingly, our approximations lead to an improved understanding of the finite and large sample properties of semiparametric estimators.

Relative to conventional approximations, the distinguishing feature of the distributional approximations developed herein is that they explicitly account for the presence of a (possibly) first-order bias effect, which emerges when the precision of the first-step nonparametric estimator is sufficiently low. The presence of the bias unearthed by our first main result poses potentially serious challenges for inference: for instance, the commonly used "estimator $\pm$ 1.96 $\times$ standard error" approach to construct an approximate 95% confidence interval for a scalar parameter of interest is invalid in the presence of a nonnegligible bias. Nonetheless, our second main result shows that a carefully implemented nonparametric bootstrap distributional approximation provides an automatic method of bias correction and that the associated percentile confidence intervals are asymptotically valid even in the presence of a nonnegligible bias. In addition to being of theoretical interest, this result therefore offers guidance for empirical work.

For the semiparametric estimators we consider, the precision of the nonparametric ingredient is governed by the bandwidth associated with the kernel-based first-step estimator. In the development of our results, we use this bandwidth as a technical device to shed light on the interplay between the distributional properties of the semiparametric estimator and the precision of its nonparametric ingredient. In particular, because the rate of convergence of the nonparametric ingredient is low when the bandwidth is "small," the bandwidths for which our results offer new insights are those that are small and we therefore use the term "small bandwidth asymptotics" to highlight the distinguishing feature of the technical approach we take in this paper. This terminology is consistent with that used in earlier work of ours, but in important respects the results obtained herein differ from those currently available in the literature.

Cattaneo, Crump, and Jansson (2010, 2014a) studied the density-weighted average derivative estimator of Powell, Stock, and Stoker (1989) and showed that the distinguishing feature emerging from the small bandwidth distributional approximation for that particular estimator is the presence of a variance effect, while Cattaneo, Crump, and Jansson (2014b) showed that the variance effect in question cannot be corrected for by using the standard nonparametric bootstrap. In contrast, this paper is concerned with a class of estimators for which the distinguishing feature of their small bandwidth asymptotic distribution is the presence of a bias effect. A well-known member of the class of estimators studied in this paper is the weighted average derivative estimator analyzed in Cattaneo, Crump, and Jansson (2013) and, as a consequence, our first main result can be interpreted as a nontrivial generalization of one of the results in that paper, since the results herein cover a large class of two-step (possibly over-identified and non-differentiable) GMM settings. Furthermore, our second main result offering bootstrap-based automatic bias reduction and valid inference appears to be new in the literature.

At a conceptual level, our small bandwidth approach is very similar to the "dimension asymptotics" approach taken in the seminal work of Mammen (1989) and, although the technical details are rather different, some of our main conclusions are similar to his. For a more detailed explanation of the connection between small bandwidth asymptotics and dimension asymptotics, see Enno Mammen's discussion of Cattaneo, Crump, and Jansson (2013). The approach we take is also similar to the approach taken by Abadie and Imbens (2006, 2008), but our main conclusion regarding the bootstrap (and subsampling) is quite different from that of Abadie and Imbens (2008).

The literature on two-step semiparametric estimators is vast, but our first main result differs from most existing results in at least two respects. First, due to the presence of a bias, our distributional conclusions differ from those obtained in the work surveyed by Andrews (1994b), Newey and McFadden (1994), Chen (2007), and Ichimura and Todd (2007). Second, a seemingly novel technical feature of our work is that reliance on a heretofore ubiquitous stochastic equicontinuity condition is avoided and that avoiding such condition is necessary, in general, in order for the bias we highlight to be nonnegligible; that is, our generalization of existing distributional conclusions cannot be accomplished without avoiding reliance on a stochastic equicontinuity condition that has featured prominently in earlier work.

Our second main result concerns the bootstrap. Previous work on bootstrap validity for general classes of semiparametric models under standard conditions includes Chen, Linton, and van Keilegom (2003) and Cheng and Huang (2010). Our result is qualitatively similar to the bootstrap consistency results of these papers, but in at least two respects our results broaden the scope of resampling-based inference in a possibly surprising way. First, we show that some (but not all) standard bootstrap-based distributional approximations deliver an automatic bias correction. Second, whereas all previous bootstrap consistency results have been obtained for settings in which subsampling-based inference procedures are also valid, the bias effect that is central to our work turns out to render subsampling-based inference procedures invalid in general. To the extent that subsampling can be regarded as a "regularized" version of the bootstrap (e.g., Bickel and Li (2006)), it therefore seems surprising that the standard nonparametric bootstrap in its simplest form turns out to be asymptotically valid in the setting of this paper.

Other work related to ours includes Chernozhukov, Escanciano, Ichimura, and Newey (2016) and Robins, Li, Tchetgen, and van der Vaart (2008). When specialized to kernel-based estimators, the local robustness property discussed by Chernozhukov et al. (2016) can be interpreted as an application of "large bandwidth asymptotics" and their results are complementary to ours in the sense that they ensure robustness to "large" bandwidths by paying more careful attention to the smoothing bias that our theory is largely silent about. The work on higher-order influence functions by Robins et al. (2008) is similar to ours at least insofar as it uses higher-order U-statistics and focuses on settings where nonparametric ingredients converge at slow rates, but unlike us they focused on problems for which optimal interval estimates exhibit a slower-than-usual rate of convergence, and even when specialized to the average density example studied below, the results obtained using their approach (e.g., Robins, Li, Tchetgen, and van der Vaart (2016), Robins, Li, Mukherjee, Tchetgen, and van derVaart (2017)) appear to be quite different from ours.

The paper proceeds as follows. Section 2 introduces the setup and gives our first main result. Section 3 gives an in-depth discussion of that result, including both connections to previous theoretical work on semiparametrics and implications for empirical work employing semiparametric inference procedures. Section 4 presents our second main result, a bootstrap analog of the main result from Section 2. Section 5 is concerned with generic verification of the high-level assumptions under which our main results are obtained, while Section 6 illustrates how the latter sufficient conditions for our high-level assumptions can be verified in the context of the specific example of inverse probability weighting (IPW) estimation with possibly non-differentiable moment functions. Finally, Section 7 offers simulation evidence, and Section 8 concludes.

Three distinct examples are considered in the paper. The first of these is mainly pedagogical and serves the dual purposes of illustrating our main results in a canonical setting while at the same time demonstrating the fact that the complications we highlight are

present even in the simplest of examples. Our second example, the IPW example already mentioned, is more substantive and a representative member of a class of estimators which is very popular in a variety of settings in applied work, including program evaluation, missing data, measurement error, and data combination. Finally, the simulation results make use of an estimator which is easy to compute, yet somewhat challenging to analyze and base inference on, namely, a so-called "Hit Rate" estimator. Technical details for all three examples are provided in the Supplemental Material (Cattaneo and Jansson (2018)), which also contains some additional technical results that may be of independent interest.

2. KERNEL-BASED SEMIPARAMETRIC ESTIMATORS

Suppose $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ is an estimand representable as the solution (with respect to $\theta \in \Theta$) to an equation of the form

$$G(\theta, \gamma_0) = 0, \qquad G(\theta, \gamma) = \mathbb{E}g(z, \theta, \gamma),$$

where g is a known functional, z is a random vector, and γ_0 is an unknown function. Letting $z_1, \ldots, z_n$ denote i.i.d. copies of z and assuming that $\hat{\gamma}_n$ is a nonparametric estimator of γ_0, a natural estimator $\hat{\theta}_n$ of θ_0 is given by a minimizer (with respect to $\theta \in \Theta$) of

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

where $\hat{W}_n$ is some (possibly random) symmetric, positive semidefinite matrix.

Estimators of this kind, often referred to as semiparametric two-step estimators, are widely used in practice and have received considerable attention in the literature. A common feature of existing distributional results for semiparametric two-step estimators, including those surveyed by Andrews (1994b), Newey and McFadden (1994), Chen (2007), and Ichimura and Todd (2007), is that they are developed under assumptions ensuring that the limiting distribution of $\hat{\theta}_n$ depends on $\hat{\gamma}_n$ only through the estimand γ_0. To be specific, existing asymptotic results are of the form

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \tag{1}$$

where $\rightsquigarrow$ denotes weak convergence and where it follows from Newey (1994a, Proposition 1), that the asymptotic variance Σ_0 depends on $\hat{\gamma}_n$ only through its probability limit (under general misspecification) and not on the method used to construct $\hat{\gamma}_n$ (e.g., kernels, local polynomials, or series) and/or on the value of the "tuning" parameter(s) associated with the chosen method (e.g., the kernel and the bandwidth for kernel estimators). While the simplicity of the limiting distribution in (1) is desirable insofar as it facilitates inference on θ_0, the rather extreme insensitivity of this distributional approximation with respect to the specifics of the nuisance parameter estimator $\hat{\gamma}_n$ is arguably unsatisfactory because folklore and simulation evidence suggests that, in samples of realistic size, the distributional properties of $\hat{\theta}_n$ do in fact depend somewhat heavily on the specifics of $\hat{\gamma}_n$.

The insensitivity of the distributional conclusion (1) with respect to the specifics of the first-step estimator $\hat{\gamma}_n$ is driven in large part by assumptions ensuring that $\hat{\gamma}_n$ converges sufficiently rapidly to γ_0. To be specific, assumptions of the form $\hat{\gamma}_n - \gamma_0 = o_{\mathbb{P}}(n^{-1/4})$ are ubiquitous in the literature on semiparametric two-step estimators, and the simplicity

of (1) is largely due to these convergence rate assumptions. As a means to the end of developing more reliable distributional approximations for $\hat{\theta}_n$, this paper allows for (but does not require) milder-than-usual convergence rate requirements on $\hat{\gamma}_n$ as a theoretical device to obtain distributional approximations for semiparametric estimators that have the intuitive appeal of featuring an explicit dependence (even asymptotically) on some of the specific features underlying the estimator $\hat{\gamma}_n$. Therefore, unlike conventional approximations currently available in the literature, our distribution theory for two-step semiparametric estimators is able to explicitly account for the effect of the first-step estimator on the distributional approximation. More specifically, we obtain results of the form

$$\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \tag{2}$$

where Σ_0 is the usual asymptotic variance of a semiparametric estimator (i.e., the same as in (1)) and $\mathfrak{B}_n$ is a non-random "bias" term. Because the distribution theory developed herein is consistent with conventional results when the latter are applicable, the bias $\mathfrak{B}_n$ in (2) is asymptotically negligible (i.e., $o(n^{-1/2})$) under conventional assumptions, but in general $\mathfrak{B}_n$ turns out to be nonnegligible under seemingly mild departures from those assumptions. Moreover, the magnitude and functional form of $\mathfrak{B}_n$ turn out to depend on the specifics of the estimator $\hat{\gamma}_n$ used in the construction of $\hat{\theta}_n$. In other words, we find that although the asymptotic variance of $\hat{\theta}_n$ remains insensitive with respect to the type of first-step nonparametric estimator also under our (weaker) assumptions, the specific structure of $\hat{\gamma}_n$ does exert a first-order effect on $\hat{\theta}_n$ through $\mathfrak{B}_n$ when milder-than-usual convergence rate requirements are placed on $\hat{\gamma}_n$.

The result (2) follows from three easy-to-interpret high-level conditions in the important special case where the first-step estimator $\hat{\gamma}_n$ is kernel-based in the sense that

$$\hat{\gamma}_n = (\hat{\gamma}_{n,1}, \ldots, \hat{\gamma}_{n,d_\gamma})', \qquad \hat{\gamma}_{n,k}(z, \theta) = \frac{1}{n}\sum_{j=1}^{n} w_k(z_j, \theta)\kappa_{n,k}\big[x_k(z, \theta) - x_k(z_j, \theta)\big], \tag{3}$$

where $\kappa_{n,k}(x) = \kappa_k(x/h_{n,k})/h_{n,k}^{d_k}$, $h_{n,k} = o(1)$ is a bandwidth, κ_k is a (kernel-like) function, and w_k and x_k are known functions of dimensions 1 and d_k, respectively. Nonparametric estimators that can be written in the form (3) include kernel estimators (e.g., of the form discussed by Newey and McFadden (1994, Section 8.3)) and local polynomial regression estimators (e.g., Fan and Gijbels (1997)). On the other hand, series estimators are not of this form, and we therefore use the term "kernel-based" when referring to the estimator in (3).

Our first high-level condition is the following.

CONDITION AL—Approximate Linearity: For some non-random $\mathcal{J}_n$ and $\mathcal{J}_0$, $\mathcal{J}_n \to \mathcal{J}_0$ and

$$\hat{\theta}_n - \theta_0 = \mathcal{J}_n \hat{G}_n(\theta_0, \hat{\gamma}_n) + o_{\mathbb{P}}\big(n^{-1/2}\big).$$

Condition AL is referred to as "approximate linearity" in recognition of the fact that the condition effectively approximates $\hat{G}_n(\theta, \gamma)$ with a function that is linear/affine with respect to θ. In particular, Condition AL is simply a representation, the displayed equality holding with $\mathcal{J}_n = \mathcal{J}_0 = I_{d_\theta}$ and without any $o_{\mathbb{P}}(n^{-1/2})$ term, in the important special case where $g(z, \theta, \gamma) = g(z, 0, \gamma) - \theta$ and $\hat{\theta}_n$ is defined as the solution to $\hat{G}_n(\theta, \hat{\gamma}_n) = 0$. More

generally, standard heuristics suggest that, under suitable regularity conditions, Condition AL will hold with $\mathcal{J}_n = \mathcal{J}_0 = -(\dot{G}_0' W_0 \dot{G}_0)^{-1} \dot{G}_0' W_0$, where $\dot{G}_0 = \partial G(\theta, \gamma_0)/\partial \theta'|_{\theta=\theta_0}$ and where W_0 is the probability limit of $\hat{W}_n$. Lemma 1 below gives conditions under which these heuristics can be made rigorous also when $\hat{\gamma}_n$ exhibits a slower-than-usual rate of convergence.

Under Condition AL, the large sample properties of $\hat{\theta}_n$ are governed by

$$\hat{G}_n(\theta_0, \hat{\gamma}_n) = \frac{1}{n}\sum_{i=1}^{n} g_0(z_i, \hat{\gamma}_n), \qquad g_0(z, \gamma) = g(z, \theta_0, \gamma).$$

Analyzing this object without assuming a faster-than-$n^{1/4}$ rate of convergence on the part of $\hat{\gamma}_n$ turns out to be challenging partly because the standard method of accounting for the dependence/overlap between the arguments z_i and $\hat{\gamma}_n$ of the summand $g_0(z_i, \hat{\gamma}_n)$ turns out to be invalid when $\hat{\gamma}_n$ converges at a slower-than-usual rate. Specifically, as further discussed and exemplified in Section 3.1, it turns out that a commonly employed stochastic equicontinuity condition typically requires (and/or is applicable only when one assumes) that the rate of convergence of $\hat{\gamma}_n$ exceeds $n^{1/4}$.

Analyzing $\hat{G}_n(\theta_0, \hat{\gamma}_n)$ without imposing further structure on g and/or relying on stochastic equicontinuity nevertheless turns out to be feasible when $\hat{\gamma}_n$ is kernel-based, the reason being that in this case $\hat{G}_n(\theta_0, \hat{\gamma}_n)$ admits a representation of the form

$$\hat{G}_n(\theta_0, \hat{\gamma}_n) = \frac{1}{n}\sum_{i=1}^{n} g_n\big(z_i, \hat{\gamma}_n^{(i)}\big), \tag{4}$$

where g_n is some function and where

$$\hat{\gamma}_n^{(i)} = \big(\hat{\gamma}_{n,1}^{(i)}, \ldots, \hat{\gamma}_{n,d_\gamma}^{(i)}\big)', \qquad \hat{\gamma}_{n,k}^{(i)}(z, \theta) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} w_k(z_j, \theta)\kappa_{n,k}\big[x_k(z, \theta) - x_k(z_j, \theta)\big],$$

is the ith "leave-one-out" estimator of γ_0. To be specific, the fact that $\hat{\gamma}_n$ is kernel-based implies that each $\hat{\gamma}_{n,k}$ is additively separable between z_i and $\{z_j : j \neq i\}$:

$$\hat{\gamma}_{n,k}(z, \theta) = n^{-1}\hat{\gamma}_{n,k}^{i}(z, \theta) + \big(1 - n^{-1}\big)\hat{\gamma}_{n,k}^{(i)}(z, \theta),$$

where

$$\hat{\gamma}_n^{i} = \big(\hat{\gamma}_{n,1}^{i}, \ldots, \hat{\gamma}_{n,d_\gamma}^{i}\big)', \qquad \hat{\gamma}_{n,k}^{i}(z, \theta) = w_k(z_i, \theta)\kappa_{n,k}\big[x_k(z, \theta) - x_k(z_i, \theta)\big].$$

As a consequence, the function

$$g_n(z_i, \gamma) = g_0\big(z_i, n^{-1}\hat{\gamma}_n^{i} + \big(1 - n^{-1}\big)\gamma\big)$$

satisfies $g_n(z_i, \hat{\gamma}_n^{(i)}) = g_0(z_i, \hat{\gamma}_n)$, implying in particular that the representation (4) is valid.

In addition to delivering (4), the assumption that $\hat{\gamma}_n$ is kernel-based makes it possible to formulate primitive conditions under which the following high-level assumption is satisfied.

CONDITION AS—Asymptotic Separability: For some $\bar{g}_n$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigl[g_n\bigl(z_i,\hat{\gamma}_n^{(i)}\bigr)-g_n(z_i,\gamma_n)\bigr]\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigl[\bar{g}_n\bigl(z_i,\hat{\gamma}_n^{(i)}\bigr)-\bar{g}_n(z_i,\gamma_n)\bigr]+o_{\mathbb{P}}(1)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigl[\bar{G}_n\bigl(\hat{\gamma}_n^{(i)}\bigr)-\bar{G}_n(\gamma_n)\bigr]+o_{\mathbb{P}}(1),
\end{aligned}$$

where $\gamma_n(\cdot)=\mathbb{E}\hat{\gamma}_n(\cdot)$ and $\bar{G}_n(\gamma)=\mathbb{E}\bar{g}_n(z,\gamma)$.

The main part of Condition AS is the second equality and the function $\bar{g}_n$ is introduced to facilitate verification of that part (and of Condition AN below). Indeed, while the first part of Condition AS holds (without any $o_{\mathbb{P}}(1)$ term) when $\bar{g}_n=g_n$, the second part of Condition AS is considerably easier to verify when $\bar{g}_n(z,\cdot)$ is a low-order polynomial approximation to $g_n(z,\cdot)$. When the rate of convergence of $\hat{\gamma}_n$ exceeds $n^{1/6}$ (but not necessarily $n^{1/4}$), the simplest polynomial approximation to $g_n(z,\cdot)$ satisfying the first part of Condition AS is usually a quadratic one of the form

$$\bar{g}_n(z,\gamma)=g_n(z,\gamma_n)+g_{n,\gamma}(z)[\gamma-\gamma_n]+\frac{1}{2}g_{n,\gamma\gamma}(z)[\gamma-\gamma_n,\gamma-\gamma_n],\tag{5}$$

where $g_{n,\gamma}(z)[\cdot]$ and $g_{n,\gamma\gamma}(z)[\cdot,\cdot]$ are linear and bilinear functionals, respectively. Conditions under which the second part of Condition AS is satisfied when $\bar{g}_n$ is of the form (5) will be given in Lemma 2 below.

Condition AS implies that the separable (between z_i and $\hat{\gamma}_n^{(i)}$) approximation

$$g_n\bigl(z_i,\hat{\gamma}_n^{(i)}\bigr)\approx g_n(z_i,\gamma_n)+\bar{G}_n\bigl(\hat{\gamma}_n^{(i)}\bigr)-\bar{G}_n(\gamma_n)$$

to $g_n(z_i,\hat{\gamma}_n^{(i)})$ is asymptotically valid in the sense that it satisfies

$$\begin{aligned}
\sqrt{n}\hat{G}_n(\theta_0,\hat{\gamma}_n)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_n\bigl(z_i,\hat{\gamma}_n^{(i)}\bigr)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bigl[g_n(z_i,\gamma_n)+\bar{G}_n\bigl(\hat{\gamma}_n^{(i)}\bigr)-\bar{G}_n(\gamma_n)\bigr]+o_{\mathbb{P}}(1).
\end{aligned}\tag{6}$$

Because averages of terms (such as $g_n(z_i,\gamma_n)$ and $\bar{G}_n(\hat{\gamma}_n^{(i)})-\bar{G}_n(\gamma_n)$) that each depend on one, but not both, of z_i and $\hat{\gamma}_n^{(i)}$ are much easier to analyze than averages of terms (such as $g_n(z_i,\hat{\gamma}_n^{(i)})$) that depend on both z_i and $\hat{\gamma}_n^{(i)}$, Condition AS therefore greatly simplifies the analysis of $\hat{G}_n(\theta_0,\hat{\gamma}_n)$.

In addition to the notational nuisance of having to employ additional sub- and superscripts in many places, a more substantive complication that must be addressed when proceeding under Condition AS is that it turns out that the leading term in (6) has a nonnegligible mean in general. Whereas the limiting distribution of $\sqrt{n}\hat{G}_n(\theta_0,\hat{\gamma}_n)$ is normal

with mean zero under conventional asymptotics, the simplest asymptotic normality result about the leading term in (6) that one can hope for more generally is therefore the following, primitive sufficient conditions for which will be given in Lemma 3 below.

CONDITION AN—Asymptotic Normality: For some non-random $\mathcal{B}_n$ and Ω_0,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g_n(z_i,\gamma_n)+\bar{G}_n\big(\hat{\gamma}_n^{(i)}\big)-\bar{G}_n(\gamma_n)-\mathcal{B}_n\right]\rightsquigarrow\mathcal{N}(0,\Omega_0).$$

Combining Conditions AL, AS, and AN, we obtain (2). For later reference, we state this observation as a theorem.

THEOREM 1: *If $\hat{\gamma}_n$ is kernel-based and if Conditions AL, AS, and AN are satisfied, then (2) holds with $\Sigma_0=\mathcal{J}_0\Omega_0\mathcal{J}_0'$ and $\mathfrak{B}_n=\mathcal{J}_n\mathcal{B}_n$.*

3. DISCUSSION OF THEOREM 1

Theorem 1 differs in three important ways from existing "master theorems" concerning the asymptotic distribution of semiparametric two-step estimators. First, although the high-level assumptions of Theorem 1 look remarkably similar to their natural counterparts in the existing literature, our Assumption AS differs in a subtle, yet crucial, way from a heretofore ubiquitous stochastic equicontinuity assumption. Second, Theorem 1 sheds new light on the bias properties of semiparametric two-step estimators. Finally, and perhaps most interestingly from the perspective of empirical practice, Theorem 1 has important implications for inference. The following subsections discuss these three differences in turn and illustrate them by means of the following canonical example.

EXAMPLE 1—Average Density: Suppose $z_1,\dots,z_n$ are i.i.d. copies of a continuously distributed random vector $z\in\mathbb{R}^d$ with a density γ_0. Then a kernel-based estimator of $\theta_0=\mathbb{E}\gamma_0(z)$, the average density, is given by

$$\hat{\theta}_n^{\mathrm{AD}}=\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_n(z_i),\qquad \hat{\gamma}_n(z)=\frac{1}{n}\sum_{j=1}^{n}K_n(z-z_j),$$

where $K_n(z)=K(z/h_n)/h_n^d$, h_n is a bandwidth, and K is a kernel. The estimator $\hat{\theta}_n^{\mathrm{AD}}$ can be interpreted as the solution to $\hat{G}_n(\theta,\hat{\gamma}_n)=0$, where

$$g(z,\theta,\gamma)=g^{\mathrm{AD}}(z,\theta,\gamma)=\gamma(z)-\theta.$$

Under standard regularity conditions (e.g., those given in Section SA.1 of the Supplemental Material), $\hat{\theta}_n^{\mathrm{AD}}$ can be analyzed using the results of this paper, as can the related estimators $\hat{\theta}_n^{\mathrm{ISD}}$ and $\hat{\theta}_n^{\mathrm{LR}}$ introduced below.

3.1. *Asymptotics Without Stochastic Equicontinuity*

In the existing semiparametrics literature, the analysis of objects such as $\hat{G}_n(\theta_0,\hat{\gamma}_n)$ invariably proceeds under an assumption of the following kind.

CONDITION SE—Stochastic Equicontinuity: For some $\bar{g}_0$,

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

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

Like Condition AS, Condition SE is an "asymptotic separability" condition insofar as it implies that the separable (between z_i and $\hat{\gamma}_n$) approximation

$$g_0(z_i,\hat{\gamma}_n)\approx g_0(z_i,\gamma_0)+\bar{G}_0(\hat{\gamma}_n)-\bar{G}_0(\gamma_0)$$

to $g_0(z_i,\hat{\gamma}_n)$ is asymptotically valid in the sense that

$$\sqrt{n}\hat{G}_n(\theta_0,\hat{\gamma}_n)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_0(z_i,\hat{\gamma}_n)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_0(z_i,\gamma_0)+\bar{G}_0(\hat{\gamma}_n)-\bar{G}_0(\gamma_0)\big]+o_{\mathbb{P}}(1).$$

We refer to the condition using the label "SE" because the second (and main) part of the condition reduces to well-known stochastic equicontinuity conditions for suitable choices of $\bar{g}_0$. In particular, the second part of Condition SE reduces to Assumption 5.2 of Newey (1994a) when $\bar{g}_0(z,\gamma)$ is linear in γ and to (2.8) of Andrews (1994a) and (3.34) of Andrews (1994b) when $\bar{g}_0=g_0$.

On the surface, Condition AS might appear to be nothing more than a "leave-one-out" counterpart of Condition SE. Crucially, however, the primitive conditions required to verify the second parts of AS and SE can often differ significantly.

EXAMPLE 1—continued: Turning first to Condition AS and setting $\bar{g}_n^{\text{AD}}=g_n^{\text{AD}}$, the first part of that condition is automatically satisfied and the second part becomes

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\hat{\gamma}_n^{(i)}(z_i)-2\gamma_n(z_i)+\theta_n\big]=o_{\mathbb{P}}(1),$$

where

$$\hat{\gamma}_n^{(i)}(z)=\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}K_n(z-z_j),\qquad \gamma_n(\cdot)=\mathbb{E}\hat{\gamma}_n(\cdot),\qquad \theta_n=\mathbb{E}\gamma_n(z).$$

It follows from a simple variance calculation that Condition AS is satisfied if $nh_n^d\to\infty$.

On the other hand, setting $\bar{g}_0^{\text{AD}}=g_0^{\text{AD}}$, the first part of Condition SE is automatically satisfied and the second part becomes

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\hat{\gamma}_n(z_i)-\gamma_n(z_i)-\gamma_0(z_i)+\theta_0\big]=o_{\mathbb{P}}(1).$$

It follows from a direct calculation that if $nh_n^d \to \infty$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\hat{\gamma}_n(z_i) - \gamma_n(z_i) - \gamma_0(z_i) + \theta_0\big] = \frac{1}{\sqrt{nh_n^{2d}}}K(0) + o_{\mathbb{P}}(1),$$

so Condition SE requires the stronger condition $nh_n^{2d} \to \infty$ unless $K(0) = 0$.

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

Although exceedingly simple in some respects, the average density example is representative in the sense that while the second part of Condition AS typically holds whenever $\hat{\gamma}_n$ is consistent (in a suitable sense), the second part of Condition SE typically requires $\hat{\gamma}_n$ to be faster-than-$n^{1/4}$-consistent. As a consequence, reliance on Condition SE must be avoided, in general, when accommodating nonparametric components whose convergence rate is no faster than $n^{1/4}$. More importantly, perhaps, the average density example illustrates the fact that reliance on Condition SE must be avoided, in general, when the goal is to generalize (1), as the term $K(0)/\sqrt{nh_n^d}$ quantifying the departure from Condition SE turns out to be the main source of the bias of the average density estimator.

In other words, in addition to being an interesting technical challenge that can be motivated by the desire to accommodate nonparametric components whose convergence rate is no faster than $n^{1/4}$, relaxing Condition SE is of fundamental importance when the goal is to obtain more refined distributional approximations than (1). 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 Condition SE (or the like) when generalizing (1) and/or accommodating nonparametric components whose convergence rate is no faster than $n^{1/4}$. Our proposed Condition AS is arguably an attractive alternative to Condition SE because it inherits one of the main benefits of the conventional Condition SE (namely, "asymptotic separability") without imposing unduly strong convergence rate requirements on $\hat{\gamma}_n$. A drawback of Condition AS in its present formulation is that $\hat{\gamma}_n$ is assumed to be kernel-based. Although doing so is beyond the scope of the present paper, it would be of interest to relax that assumption.

We are aware of only two exceptions to the rule that Condition SE requires $\hat{\gamma}_n$ to be faster-than-$n^{1/4}$-consistent. The first of these exceptions occurs when $g_0(z_i, \gamma)$ and $g_n(z_i, \gamma)$ coincide (apart from a non-important factor of proportionality). An important example of this phenomenon is provided by the "leave-in" version of Powell, Stock, and Stoker's (1989) estimator: As pointed out in their footnote 6, that estimator satisfies $g_0(z_i, \gamma) = (1 - n^{-1})g_n(z_i, \gamma)$ because symmetric kernels satisfy $K'(0) = 0$. The other exception occurs when $g_0(z, \gamma)$ is already additively separable between z and γ, as is the case for the consumer surplus estimator of Hausman and Newey (1995) where the associated $g_0(z, \gamma)$ does not depend on z at all. Both exceptions can be illustrated by means of Example 1.

EXAMPLE 1—continued: The function g_0^{AD} satisfies $g_0^{\text{AD}}(z_i, \gamma) = (1 - n^{-1})g_n^{\text{AD}}(z_i, \gamma)$ when $K(0) = 0$, so in this case Condition SE holds whenever Condition AS does.

An alternative estimator of $\theta_0 = \int_{\mathbb{R}^d} \gamma_0(u)^2\,du$ is the integrated squared density estimator

$$\hat{\theta}_n^{\mathrm{ISD}} = \int_{\mathbb{R}^d} \hat{\gamma}_n(u)^2\,du,$$

which can be interpreted as the solution to $\hat{G}_n(\theta, \hat{\gamma}_n) = 0$, where

$$g(z, \theta, \gamma) = g^{\mathrm{ISD}}(z, \theta, \gamma) = \int_{\mathbb{R}^d} \gamma(u)^2\,du - \theta.$$

Because $g_0^{\mathrm{ISD}}(z, \gamma) = \int_{\mathbb{R}^d} \gamma(u)^2\,du - \theta_0$ does not even depend on z, (asymptotic) "separability" between z and γ is of course automatic and, indeed, both parts of Condition SE are satisfied (without any $o_{\mathbb{P}}(1)$ terms) when $\bar{g}_0^{\mathrm{ISD}} = g_0^{\mathrm{ISD}}$. (Setting $\bar{g}_n^{\mathrm{ISD}} = g_n^{\mathrm{ISD}}$ and applying Lemma 2 below, Condition AS can also be shown to hold provided $nh_n^d \to \infty$.)

3.2. *Bias Properties*

Under the conditions of Theorem 1, the main determinant of the bias $\mathfrak{B}_n$ in (2) is $\mathcal{B}_n$ of Condition AN. When Condition AS is satisfied with a $\bar{g}_n$ of the form (5), the functional $\bar{G}_n$ is also quadratic. Indeed, defining

$$G_n(\gamma) = \mathbb{E}g_n(z, \gamma), \qquad G_{n,\gamma}[\eta] = \mathbb{E}g_{n,\gamma}(z)[\eta], \qquad G_{n,\gamma\gamma}[\eta, \varphi] = \mathbb{E}g_{n,\gamma\gamma}(z)[\eta, \varphi],$$

we have

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

Because $\hat{\gamma}_{i,n} - \gamma_n$ has mean zero, the leading term in (6) therefore satisfies

$$\mathbb{E}\big[g_n(z_i, \gamma_n) + \bar{G}_n\big(\hat{\gamma}_n^{(i)}\big) - \bar{G}_n(\gamma_n)\big] = \mathcal{B}_n^{\mathrm{S}} + \mathcal{B}_n^{\mathrm{LI}} + \mathcal{B}_n^{\mathrm{NL}},$$

where

$$\mathcal{B}_n^{\mathrm{S}} = G_0(\gamma_n), \qquad G_0(\gamma) = \mathbb{E}g_0(z, \gamma),$$

is a "smoothing" bias term, while

$$\mathcal{B}_n^{\mathrm{LI}} = G_n(\gamma_n) - G_0(\gamma_n) \quad \text{and} \quad \mathcal{B}_n^{\mathrm{NL}} = \frac{1}{2n}\mathbb{E}G_{n,\gamma\gamma}\big[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n\big]$$

are generic versions of what Cattaneo, Crump, and Jansson (2013) referred to as "leave-in" and "nonlinearity" bias terms, respectively.

The smoothing bias $\mathcal{B}_n^{\mathrm{S}}$ is familiar from the conventional theory and we have nothing new to say about it, but because one of our main results (namely, Theorem 2 below) effectively requires the smoothing bias to be asymptotically negligible (i.e., $\mathcal{B}_n^{\mathrm{S}} = o(n^{-1/2})$), we give a brief discussion of sufficient conditions for this to occur. In most cases, the magnitude of $\mathcal{B}_n^{\mathrm{S}}$ coincides with that of the smoothing bias $\gamma_n - \gamma_0$ of the first-step estimator $\hat{\gamma}_n$, leading to the familiar conclusion that undersmoothing is required in order to achieve $\mathcal{B}_n^{\mathrm{S}} = o(n^{-1/2})$. An exception to this rule might occur when the moment function $g(z, \theta, \gamma)$ is "locally robust" in the sense of Chernozhukov et al. (2016), as $\hat{\theta}_n$ then has the "small bias property" discussed by Newey, Hsieh, and Robins (2004); that is, the magnitude of $\mathcal{B}_n^{\mathrm{S}}$ is smaller than that of $\gamma_n - \gamma_0$.

EXAMPLE 1—continued: The bias $\gamma_n - \gamma_0$ of $\hat{\gamma}_n$ satisfies $\int_{\mathbb{R}^d}[\gamma_n(u) - \gamma_0(u)]^2\,du = O(h_n^{2P})$, as $h_n \to 0$, where P is the order of the kernel K. As a consequence,

$$G_0^{\mathrm{AD}}(\gamma_n) = \int_{\mathbb{R}^d}\big[\gamma_n(u) - \gamma_0(u)\big]\gamma_0(u)\,du = O\big(h_n^P\big),$$

so the smoothing bias associated with $\hat{\theta}_n^{\mathrm{AD}}$ is asymptotically negligible provided $nh_n^{2P} \to 0$, a condition which requires undersmoothing because the MSE-optimal bandwidth for $\hat{\gamma}_n$ satisfies $h_n \sim n^{-1/(2P+d)}$.

The condition for the smoothing bias associated with $\hat{\theta}_n^{\mathrm{ISD}}$ to be asymptotically negligible is the same as that for $\hat{\theta}_n^{\mathrm{AD}}$, the reason being that

$$G_0^{\mathrm{ISD}}(\gamma_n) = 2G_0^{\mathrm{AD}}(\gamma_n) + \int_{\mathbb{R}^d}\big[\gamma_n(u) - \gamma_0(u)\big]^2\,du = 2G_0^{\mathrm{AD}}(\gamma_n) + O\big(h_n^{2P}\big).$$

On the other hand, the estimator

$$\hat{\theta}_n^{\mathrm{LR}} = 2\hat{\theta}_n^{\mathrm{AD}} - \hat{\theta}_n^{\mathrm{ISD}} = \frac{2}{n}\sum_{i=1}^n \hat{\gamma}_n(z_i) - \int_{\mathbb{R}^d}\hat{\gamma}_n(u)^2\,du$$

has the small bias property, as it can be interpreted as the solution to $\hat{G}_n(\theta, \hat{\gamma}_n) = 0$ with

$$g(z, \theta, \gamma) = g^{\mathrm{LR}}(z, \theta, \gamma) = 2g^{\mathrm{AD}}(z, \theta, \gamma) - g^{\mathrm{ISD}}(z, \theta, \gamma) = 2\gamma(z) - \int_{\mathbb{R}^d}\gamma(u)^2\,du - \theta,$$

where g^{LR} is locally robust because it follows from the foregoing that

$$G_0^{\mathrm{LR}}(\gamma_n) = -\int_{\mathbb{R}^d}\big[\gamma_n(u) - \gamma_0(u)\big]^2\,du = O\big(h_n^{2P}\big).$$

As a consequence, the smoothing bias associated with $\hat{\theta}_n^{\mathrm{LR}}$ is asymptotically negligible provided $nh_n^{4P} \to 0$, a condition which does not require undersmoothing when $P > d/2$.

The leave-in and nonlinearity biases are usually asymptotically negligible whenever the rate of convergence of $\hat{\gamma}_n$ exceeds $n^{1/4}$. As a consequence, these biases play no role in the conventional theory. In contrast, it turns out that one or both of $\mathcal{B}_n^{\mathrm{LI}}$ and $\mathcal{B}_n^{\mathrm{NL}}$ will typically be nonnegligible when the rate of convergence of $\hat{\gamma}_n$ is no faster than $n^{1/4}$. To be specific, when $\hat{\gamma}_n - \gamma_n \neq o_{\mathbb{P}}(n^{1/4})$, one typically finds that $\mathcal{B}_n^{\mathrm{LI}}$ is nonnegligible whenever Condition SE fails while $\mathcal{B}_n^{\mathrm{NL}}$ is nonnegligible whenever $g_0(z, \gamma)$ is nonlinear in γ.

EXAMPLE 1—continued: Because

$$G_n^{\mathrm{AD}}(\gamma_n) - G_0^{\mathrm{AD}}(\gamma_n) = \frac{1}{nh_n^d}K(0) + O\big(n^{-1}\big),$$

the leave-in bias associated with $\hat{\theta}_n^{\mathrm{AD}}$ is nonnegligible unless either $nh_n^{2d} \to \infty$ or $K(0) = 0$, the former being the condition under which the rate of convergence of $\hat{\gamma}_n$ exceeds $n^{1/4}$ and the latter being the condition under which Condition SE is satisfied by g^{AD}. On the other hand, because $g_0^{\mathrm{AD}}(z, \gamma) = \gamma(z) - \theta_0$ is linear in γ, $G_{n,\gamma\gamma}^{\mathrm{AD}}[\cdot, \cdot] = 0$ and the nonlinearity

bias associated with $\hat{\theta}_n^{\mathrm{AD}}$ is zero. In summary, we therefore find that if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then

$$\mathbb{E}\big[g_n^{\mathrm{AD}}(z_i, \gamma_n) + \bar{G}_n^{\mathrm{AD}}\big(\hat{\gamma}_n^{(i)}\big) - \bar{G}_n^{\mathrm{AD}}(\gamma_n)\big] = \mathfrak{B}_n^{\mathrm{AD}} + o\big(n^{-1/2}\big), \qquad \mathfrak{B}_n^{\mathrm{AD}} = \frac{1}{nh_n^d} K(0).$$

When $nh_n^d \to \infty$, Condition SE is satisfied by g^{ISD} and the leave-in bias associated with $\hat{\theta}_n^{\mathrm{ISD}}$ is negligible because

$$G_n^{\mathrm{ISD}}(\gamma_n) - G_0^{\mathrm{ISD}}(\gamma_n) = O\big(n^{-1}\big).$$

On the other hand, because $g_0^{\mathrm{ISD}}(z, \gamma) = \int_{\mathbb{R}^d} \gamma(u)^2\, du - \theta_0$ is nonlinear in γ, the nonlinearity bias associated with $\hat{\theta}_n^{\mathrm{ISD}}$ is nonzero. Indeed,

$$\mathbb{E} G_{n,\gamma\gamma}^{\mathrm{ISD}}\big[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n\big] = \frac{2}{h_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n)\, du\, dv + O\big(1 + n^{-1} h_n^{-d}\big),$$

so the nonlinearity bias associated with $\hat{\theta}_n^{\mathrm{ISD}}$ is nonnegligible unless $nh_n^{2d} \to \infty$. In summary, we therefore find that if $nh_n^{2P} \to 0$ and if $nh_n^d \to \infty$, then

$$\mathbb{E}\big[g_n^{\mathrm{ISD}}(z_i, \gamma_n) + \bar{G}_n^{\mathrm{ISD}}\big(\hat{\gamma}_n^{(i)}\big) - \bar{G}_n^{\mathrm{ISD}}(\gamma_n)\big] = \mathfrak{B}_n^{\mathrm{ISD}} + o\big(n^{-1/2}\big),$$

where

$$\mathfrak{B}_n^{\mathrm{ISD}} = \frac{1}{nh_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n)\, du\, dv.$$

Finally, being a linear combination of $\hat{\theta}_n^{\mathrm{AD}}$ and $\hat{\theta}_n^{\mathrm{ISD}}$, the locally robust estimator $\hat{\theta}_n^{\mathrm{LR}}$ has nonnegligible leave-in and nonlinearity biases associated with it unless $nh_n^{2d} \to \infty$. To be specific, it follows from the foregoing that if $nh_n^{4P} \to 0$ and if $nh_n^d \to \infty$, then

$$\mathbb{E}\big[g_n^{\mathrm{LR}}(z_i, \gamma_n) + \bar{G}_n^{\mathrm{LR}}\big(\hat{\gamma}_n^{(i)}\big) - \bar{G}_n^{\mathrm{LR}}(\gamma_n)\big] = \mathfrak{B}_n^{\mathrm{LR}} + o\big(n^{-1/2}\big),$$

where

$$\mathfrak{B}_n^{\mathrm{LR}} = \frac{1}{nh_n^d}\left[2K(0) - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(v)^2 \gamma_0(u - vh_n)\, du\, dv\right].$$

3.3. *Inference*

Because (2) generalizes to the familiar result (1) by accommodating $\mathfrak{B}_n \neq 0$, it is natural to investigate whether inference procedures designed to be valid under (1) remain valid also when $\mathfrak{B}_n \neq 0$ in (2). For the purposes of that investigation, the remainder of this section assumes for specificity, but without loss of relevance, that $d_\theta = 1$ (i.e., that θ_0 is scalar) and that Σ_0 is positive.

When $\hat{\theta}_n$ is assumed to satisfy (1), 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_0. If $\hat{\Sigma}_n$ is consistent, then the distributional approximation is itself consistent in the sense that

$$\sup_{t \in \mathbb{R}^{d_\theta}} \big|\mathbb{P}\big[\sqrt{n}(\hat{\theta}_n - \theta_0) \leq t\big] - \mathbb{P}\big[\mathcal{N}(0, \hat{\Sigma}_n) \leq t\big]\big| = o(1),$$

a fact which in turn implies, for instance, that the asymptotic coverage probability of the following "Normal" confidence interval for θ_0 is $1-\alpha$:

$$\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha} = [\hat{\theta}_n - \hat{q}_{n,1-\alpha/2},\, \hat{\theta}_n - \hat{q}_{n,\alpha/2}],$$

where $\hat{q}_{n,\alpha} = \inf\{q \in \mathbb{R} : \mathbb{P}[\mathcal{N}(0, \hat{\Sigma}_n) \le q] \ge \alpha\} = \Phi^{-1}(\alpha)\sqrt{\hat{\Sigma}_n/n}$, with $\Phi(\cdot)$ the standard normal c.d.f. As it turns out, replacing (1) with (2) severely affects the properties of the confidence interval $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$. Indeed, if $\hat{\Sigma}_n$ is consistent and if (2) holds, then it can be shown that

$$\mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}\big] = \Phi\big[\Phi^{-1}(1-\alpha/2) - \sqrt{n}\mathfrak{B}_n/\sqrt{\Sigma_0}\big] - \Phi\big[\Phi^{-1}(\alpha/2) - \sqrt{n}\mathfrak{B}_n/\sqrt{\Sigma_0}\big] + o(1),$$

implying in particular that $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$ is asymptotically valid if and only if $\mathfrak{B}_n = o(n^{-1/2})$.

A conceptually distinct distributional approximation is provided by the bootstrap. In standard notation, the bootstrap approximation to the c.d.f. of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is given by $\mathbb{P}^*[\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) \le \cdot]$, 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, it is well understood that asymptotically valid inference procedures can be based on the bootstrap whenever the bootstrap consistency condition

$$\sup_{t\in\mathbb{R}^{d_\theta}} \big|\mathbb{P}\big[\sqrt{n}(\hat{\theta}_n - \theta_0) \le t\big] - \mathbb{P}^*\big[\sqrt{n}\big(\hat{\theta}^*_n - \hat{\theta}_n\big) \le t\big]\big| = o_{\mathbb{P}}(1) \tag{7}$$

is satisfied.

For instance, (7) ensures that certain bootstrap-based variance estimators are consistent under (1). As a consequence, a fully "automatic" (in the sense that it can be implemented without even characterizing Σ_0) version of $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$ can be constructed by basing the variance estimator on the bootstrap, but because bootstrap-based variance estimators are consistent also under (2) (when (7) holds), the corresponding interval $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$ is asymptotically invalid under (2).

Three other well-known examples of bootstrap-based confidence intervals for θ_0 with asymptotic coverage probability $1-\alpha$ under (1) and (7) are the "Efron" interval

$$\mathsf{CI}^{\mathrm{E}}_{n,1-\alpha} = \big[\hat{\theta}_n + q^*_{n,\alpha/2},\, \hat{\theta}_n + q^*_{n,1-\alpha/2}\big],$$

the "percentile" interval

$$\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha} = \big[\hat{\theta}_n - q^*_{n,1-\alpha/2},\, \hat{\theta}_n - q^*_{n,\alpha/2}\big],$$

and the "symmetric" interval

$$\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha} = \big[\hat{\theta}_n - Q^*_{n,1-\alpha},\, \hat{\theta}_n + Q^*_{n,1-\alpha}\big],$$

where $q^*_{n,\alpha} = \inf\{q \in \mathbb{R} : \mathbb{P}^*[(\hat{\theta}^*_n - \hat{\theta}_n) \le q] \ge \alpha\}$ and $Q^*_{n,\alpha} = \inf\{Q \in \mathbb{R} : \mathbb{P}^*[|\hat{\theta}^*_n - \hat{\theta}_n| \le Q] \ge \alpha\}$.

Like $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$, the interval $\mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}$ is typically asymptotically invalid under (2). Indeed, if (2) and (7) hold, then it can be shown that

$$\mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}\big] = \Phi\big[\Phi^{-1}(1-\alpha/2) - 2\sqrt{n}\mathfrak{B}_n/\sqrt{\Sigma_0}\big] - \Phi\big[\Phi^{-1}(\alpha/2) - 2\sqrt{n}\mathfrak{B}_n/\sqrt{\Sigma_0}\big] + o(1),$$

implying in particular that $\mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}$ is asymptotically invalid when $\mathcal{B}_n \neq o(n^{-1/2})$, being even more sensitive to the bias $\mathcal{B}_n$ than $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$. On the other hand, it can be shown that (2) and (7) are sufficient to guarantee asymptotic validity of the intervals $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$; that is, if (2) and (7) hold, then

$$\mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}\big] \to 1-\alpha \quad \text{and} \quad \mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}\big] \to 1-\alpha.$$

Specializing to the "knife-edge" case where $\mathcal{B}_n \sim n^{-1/2}$, our main qualitative findings can be summarized as follows.

PROPOSITION 1: *Suppose* (2) *holds with* $\mathcal{B}_n = \mathcal{B}/\sqrt{n} + o(n^{-1/2})$ *for some* $\mathcal{B} \neq 0$. *If* $\hat{\Sigma}_n \to_{\mathbb{P}} \Sigma_0$ *and if* (7) *holds, then*

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}\big] &< \lim_{n\to\infty} \mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}\big] < \lim_{n\to\infty} \mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}\big] \\ &= \lim_{n\to\infty} \mathbb{P}\big[\theta_0 \in \mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}\big] = 1-\alpha.\end{aligned}$$

The main constructive message of Proposition 1 and the discussion preceding it is that replacing (1) with (2) would not have serious consequences for the coverage probabilities of the intervals $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ if validity of (7) could be established also under (2). Conditions for this to occur are given in the next section.

Although $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ enjoy similar coverage properties, their efficiency properties can be very different. Indeed, if (2) and (7) hold, then $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ is rate-optimal in the sense that its width $q^*_{n,1-\alpha/2} - q^*_{n,\alpha/2}$ is $O_{\mathbb{P}}(n^{-1/2})$. In contrast, $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ has width $2\mathcal{Q}^*_{n,1-\alpha} = 2|\mathcal{B}_n| + O_p(n^{-1/2})$, implying in particular that it is not even rate-optimal when $\sqrt{n}|\mathcal{B}_n| \to \infty$. More generally, $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ is (asymptotically) wider than $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ whenever $\mathcal{B}_n \neq o(n^{-1/2})$.

We close this section by briefly discussing three additional types of confidence intervals that are known to be "robust" in the sense that they do not require a consistent estimator of Σ_0 or even the full force of the $\sqrt{n}$-normality property (1). First, the inference procedure of Ibragimov and Müller (2010) can be adapted to the current setup to produce a confidence interval whose asymptotic validity follows from (1) even if Σ_0 does not admit a consistent estimator. Second, in the more general case where $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a (non-degenerate) limiting distribution which is symmetric about zero, then the procedure recently proposed by Canay, Romano, and Shaikh (2017) can be used to construct an asymptotically valid confidence interval for θ_0. Finally, in the yet more general case where one makes only the "minimal" assumption that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a (non-degenerate) limiting distribution, then the subsampling approximation to the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is known to be consistent (e.g., Politis and Romano (1994)). Like $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}$, confidence intervals based on the procedures of Ibragimov and Müller (2010) and Canay, Romano, and Shaikh (2017) are asymptotically invalid if $\mathcal{B}_n \neq o(n^{-1/2})$. Subsampling-based confidence intervals, on the other hand, are valid provided $\sqrt{n}\mathcal{B}_n$ is convergent (not necessarily to zero), but even these intervals are invalid in general if $\mathcal{B}_n \neq O(n^{-1/2})$. In particular, and perhaps surprisingly in light of the fact that subsampling is often regarded as a "regularized" version of the bootstrap (e.g., Bickel and Li (2006)), one by-product of the results of this paper is a remarkably simple example of an instance where the bootstrap-based confidence intervals $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ are asymptotically valid even though subsampling-based confidence intervals are not.

EXAMPLE 1—continued: If the bandwidth is of the form $h_n = Cn^{-1/\eta}$, where $C > 0$ and $\eta \in (d, 2P)$ are user-chosen constants, then

$$\sqrt{n}\big(\hat{\theta}_n^{\mathrm{AD}} - \theta_0 - \mathfrak{B}_n^{\mathrm{AD}}\big) \rightsquigarrow \mathcal{N}(0, \Sigma_0), \quad \Sigma_0 = 4\mathbb{V}\big[\gamma_0(z)\big].$$

Unless $K(0) = 0$, asymptotic validity of the confidence intervals $\mathsf{CI}^{\mathrm{N}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{E}}_{n,1-\alpha}$ therefore fails whenever $\eta \in (d, 2d]$. The same is true for the intervals based on the procedures of Ibragimov and Müller (2010) and Canay, Romano, and Shaikh (2017). Subsampling-based confidence intervals, on the other hand, are valid when $\eta = 2d$, but even these intervals can be shown to be invalid for $\eta \in (d, 2d)$. In contrast, as further discussed below, the intervals $\mathsf{CI}^{\mathrm{P}}_{n,1-\alpha}$ and $\mathsf{CI}^{\mathrm{S}}_{n,1-\alpha}$ turn out to be valid also when $\eta \in (d, 2d)$.

Similar remarks apply to $\hat{\theta}_n^{\mathrm{ISD}}$ and $\hat{\theta}_n^{\mathrm{LR}}$, as

$$\sqrt{n}\big(\hat{\theta}_n^{\mathrm{ISD}} - \theta_0 - \mathfrak{B}_n^{\mathrm{ISD}}\big) \rightsquigarrow \mathcal{N}(0, \Sigma_0) \quad \text{and} \quad \sqrt{n}\big(\hat{\theta}_n^{\mathrm{LR}} - \theta_0 - \mathfrak{B}_n^{\mathrm{LR}}\big) \rightsquigarrow \mathcal{N}(0, \Sigma_0)$$

whenever $\eta \in (d, 2P)$ and $\eta \in (d, 4P)$, respectively.

4. BOOTSTRAP CONSISTENCY

One consequence of replacing (1) with (2) is that the statistics $\sqrt{n}(\hat{\theta}_n - \theta_0)$ might cease to be tight, as $\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}\mathfrak{B}_n + O_{\mathbb{P}}(1)$ when (2) holds. 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, suppose (2) and the following bootstrap counterpart thereof hold:

$$\sqrt{n}\big(\hat{\theta}_n^* - \hat{\theta}_n - \mathfrak{B}_n^*\big) \rightsquigarrow_{\mathbb{P}} \mathcal{N}\big(0, \Sigma_0^*\big), \tag{8}$$

where $\mathfrak{B}_n^*$ and Σ_0^* are some non-random matrices and where $\rightsquigarrow_{\mathbb{P}}$ denotes weak convergence in probability. Assuming Σ_0 is positive definite, it then follows from the relation

$$\begin{aligned}&\sup_{t \in \mathbb{R}^{d_\theta}} \big|\mathbb{P}\big[\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathfrak{B}_n) \le t\big] - \mathbb{P}^*\big[\sqrt{n}\big(\hat{\theta}_n^* - \hat{\theta}_n - \mathfrak{B}_n\big) \le t\big]\big| \\ &\quad = \sup_{t \in \mathbb{R}^{d_\theta}} \big|\mathbb{P}\big[\sqrt{n}(\hat{\theta}_n - \theta_0) \le t\big] - \mathbb{P}^*\big[\sqrt{n}\big(\hat{\theta}_n^* - \hat{\theta}_n\big) \le t\big]\big|\end{aligned}$$

that a necessary and sufficient condition for (7) is that $\mathfrak{B}_n^* = \mathfrak{B}_n + o(n^{-1/2})$ and $\Sigma_0^* = \Sigma_0$.

This characterization is very useful because it turns out that (8) can be often verified by imitating the proof of (2). To give a precise statement, let $\hat{\theta}_n^*$ be a minimizer of

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

where $z_{1,n}^*, \ldots, z_{n,n}^*$ is a random sample with replacement from $z_1, \ldots, z_n$, $\hat{W}_n^*$ is some bootstrap counterpart of $\hat{W}_n$, and where

$$\hat{\gamma}_n^* = \big(\hat{\gamma}_{n,1}^*, \ldots, \hat{\gamma}_{n,d_\gamma}^*\big)', \qquad \hat{\gamma}_{n,k}^*(z, \theta) = \frac{1}{n}\sum_{j=1}^n w_k\big(z_{j,n}^*, \theta\big)\kappa_{n,k}\big[x_k(z, \theta) - x_k\big(z_{j,n}^*, \theta\big)\big].$$

Under regularity conditions, it follows from a bootstrap counterpart of Condition AL that the large sample properties of $\hat{\theta}_n^*$ are governed by $\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*)$. Moreover, in perfect analogy with (4), the fact that $\hat{\gamma}_n^*$ is kernel-based implies that

$$\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) = \frac{1}{n}\sum_{i=1}^n g_0^*(z_{i,n}^*, \hat{\gamma}_n^*) = \frac{1}{n}\sum_{i=1}^n g_n^*(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}), \tag{9}$$

where

$$\hat{\gamma}_n^{*,(i)} = (\hat{\gamma}_{n,1}^{*,(i)}, \ldots, \hat{\gamma}_{n,d_\gamma}^{*,(i)})',$$

$$\hat{\gamma}_{n,k}^{*,(i)}(z, \theta) = \frac{1}{n-1}\sum_{j=1, j\neq i}^n w_k(z_{j,n}^*, \theta)\kappa_{n,k}[x_k(z, \theta) - x_k(z_{j,n}^*, \theta)],$$

is the ith "leave-one-out" estimator of γ_0 and where, defining

$$\hat{\gamma}_n^{*,i} = (\hat{\gamma}_{n,1}^{*,i}, \ldots, \hat{\gamma}_{n,d_\gamma}^{*,i})', \qquad \hat{\gamma}_{n,k}^{*,i}(z, \theta) = w_k(z_{i,n}^*, \theta)\kappa_{n,k}[x_k(z, \theta) - x_k(z_{i,n}^*, \theta)],$$

the functions g_n^* and g_0^* satisfy

$$g_n^*(z_{i,n}^*, \gamma) = g_0^*[z_{i,n}^*, n^{-1}\hat{\gamma}_n^{*,i} + (1 - n^{-1})\gamma], \qquad g_0^*(z, \gamma) = g(z, \hat{\theta}_n, \gamma).$$

As a consequence, $\hat{\theta}_n^*$ enjoys large sample properties analogous to those of $\hat{\theta}_n$ provided bootstrap analogues of Conditions AS and AN hold.

Theorem 2 below gives a precise statement. That statement involves the following bootstrap analogues of Conditions AL, AS, and AN.

CONDITION AL*: For some non-random $\mathcal{J}_n^*$ and $\mathcal{J}_0^*$, $\mathcal{J}_n^* \to \mathcal{J}_0^*$ and

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

CONDITION AS*: For some function $\bar{g}_n^*$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n [g_n^*(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}) - g_n^*(z_{i,n}^*, \hat{\gamma}_n)] &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [\bar{g}_n^*(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}) - \bar{g}_n^*(z_{i,n}^*, \hat{\gamma}_n)] + o_{\mathbb{P}}(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [\bar{G}_n^*(\hat{\gamma}_n^{*,(i)}) - \bar{G}_n^*(\hat{\gamma}_n)] + o_{\mathbb{P}}(1),\end{aligned}$$

where $\bar{G}_n^*(\gamma) = \mathbb{E}^*\bar{g}_n^*(z_{i,n}^*, \gamma)$ and where $\mathbb{E}^*[\cdot]$ denotes $\mathbb{E}[\cdot|z_1, \ldots, z_n]$.

CONDITION AN*: For some non-random $\mathcal{B}_n^*$ and Ω_0^*,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [g_n^*(z_{i,n}^*, \hat{\gamma}_n) + \bar{G}_n^*(\hat{\gamma}_n^{*,(i)}) - \bar{G}_n^*(\hat{\gamma}_n) - \mathcal{B}_n^*] \rightsquigarrow_{\mathbb{P}} \mathcal{N}(0, \Omega_0^*).$$

THEOREM 2: *If $\hat{\gamma}_n^*$ is kernel-based and if Conditions AL*, AS*, and AN* are satisfied, then (8) holds with $\Sigma_0^* = \mathcal{J}_0^*\Omega_0^*\mathcal{J}_0^{*\prime}$ and $\mathfrak{B}_n^* = \mathcal{J}_n^*\mathcal{B}_n^*$. In particular, (7) is satisfied if (2) holds and if $\mathfrak{B}_n^* = \mathfrak{B}_n + o(n^{-1/2})$ and $\Sigma_0^* = \Sigma_0$, where Σ_0 is positive definite.*

As further demonstrated in Section 5.4, Conditions AL*, AS*, and AN* are natural bootstrap analogues of the conditions of Theorem 1 not only in appearance, but also in the sense that they can be verified by mimicking the verification of their counterparts in Theorem 1. Moreover, in most cases the conditions for bootstrap consistency given in Theorem 2 are satisfied under conditions similar to those imposed in order to obtain (2). In particular, bootstrap consistency does not require faster-than-$n^{1/4}$-consistency on the part of $\hat{\gamma}_n$.

EXAMPLE 1—continued: If $h_n \to 0$ and if $nh_n^d \to \infty$, then $\hat{\theta}_n^{\mathrm{AD},*}$, $\hat{\theta}_n^{\mathrm{ISD},*}$, and $\hat{\theta}_n^{\mathrm{LR},*}$ all satisfy (8) with $\Sigma_0^* = 4\mathbb{V}\gamma_0(z)$ and $\mathfrak{B}_n^*$ equal to $\mathfrak{B}_n^{\mathrm{AD}}$, $\mathfrak{B}_n^{\mathrm{ISD}}$, and $\mathfrak{B}_n^{\mathrm{LR}}$, respectively. As a consequence, if the bandwidth is of the form $h_n = Cn^{-1/\eta}$, then $\hat{\theta}_n^{\mathrm{AD},*}$, $\hat{\theta}_n^{\mathrm{ISD},*}$, and $\hat{\theta}_n^{\mathrm{LR},*}$ satisfy (7) whenever $\eta \in (d, 2P)$, $\eta \in (d, 2P)$, and $\eta \in (d, 4P)$, respectively.

REMARK 1: 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 bootstrap consistency results 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 over-identified models. 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.

5. VERIFYING THE ASSUMPTIONS OF THEOREMS 1 AND 2

The purpose of this section is to present tools that can be used to verify those elements of the assumptions of Theorems 1 and 2 that have no obvious counterpart in the conventional theory on semiparametric two-step estimators.

5.1. *Condition AL*

Letting $\dot{G}(\gamma)$ denote $\partial G(\theta, \gamma)/\partial\theta'|_{\theta=\theta_0}$ whenever the derivative exists (and zero otherwise), standard heuristics suggest that under suitable regularity conditions, Condition AL will hold with $\mathcal{J}_n = \mathcal{J}_0 = -(\dot{G}_0' W_0 \dot{G}_0)^{-1}\dot{G}_0' W_0$, where $\dot{G}_0 = \dot{G}(\gamma_0)$ and where W_0 is the probability limit of $\hat{W}_n$. When $\hat{G}_n(\theta_0, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/2})$, these standard heuristics can be made rigorous with the help of Pakes and Pollard (1989, Theorem 3.3), a variant of which is given by the $\rho = 2$ version of Lemma 1 below.

However, the condition $\hat{G}_n(\theta_0, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/2})$ fails, in general, when the weaker Conditions AS and AN are used to obtain distributional approximations, so in order to justify our reliance on Condition AL it is important to have sufficient conditions for Condition AL that do not require $\hat{G}_n(\theta_0, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/2})$. This observation motivates condition (iv) of the following result, whose formulation and content is in the spirit of Pakes and Pollard (1989, Theorem 3.3).

LEMMA 1: *Suppose that* $\hat{\theta}_n - \theta_0 = o_p(1)$, *that* $\dot{G}_0' W_0 \dot{G}_0$ *has rank* d_θ, *and that, for some* $\rho \in [2, 4)$ *and for some non-random* W_n *and* $\dot{G}_n$ *with* $W_n - W_0 = o(1)$ *and* $\dot{G}_n - \dot{G}_0 = o(1)$:

(i) $\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \leq \inf_{\theta\in\Theta} \hat{G}_n(\theta, \hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\theta, \hat{\gamma}_n) + o_{\mathbb{P}}(n^{-1})$;

(ii) *for every* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \frac{\|G(\theta,\hat\gamma_n) - G(\theta_0,\hat\gamma_n) - \dot G(\hat\gamma_n)(\theta-\theta_0)\|}{\|\theta-\theta_0\|^{\rho/2}} = o_{\mathbb{P}}(1);$$

(iii) *for every* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \frac{\|\hat G_n(\theta,\hat\gamma_n) - G(\theta,\hat\gamma_n) - \hat G_n(\theta_0,\hat\gamma_n) + G(\theta_0,\hat\gamma_n)\|}{1+n^{1/\rho}\|\theta-\theta_0\|} = o_{\mathbb{P}}\big(n^{-1/\rho}\big);$$

(iv) $\hat G_n(\theta_0,\hat\gamma_n) = O_{\mathbb{P}}(n^{-1/\rho})$;
(v) θ_0 *is an interior point of* Θ;
(vi) $\hat W_n - W_n = o_{\mathbb{P}}(n^{1/\rho-1/2})$ *and* $\dot G(\hat\gamma_n) - \dot G_n = o_{\mathbb{P}}(n^{1/\rho-1/2})$;
(vii) $\dot G(\hat\gamma_n)'\hat W_n \hat G_n(\hat\theta_n,\hat\gamma_n) = o_{\mathbb{P}}(n^{-1/2})$ *and, for every* $\delta_n = O(n^{-1/\rho})$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \|\hat G_n(\theta,\hat\gamma_n) - G(\theta,\hat\gamma_n) - \hat G_n(\theta_0,\hat\gamma_n) + G(\theta_0,\hat\gamma_n)\| = o_{\mathbb{P}}\big(n^{-1/2}\big).$$

Then Condition AL *holds with*

$$\mathcal{J}_n = -\big(\dot G_n' W_n \dot G_n\big)^{-1}\dot G_n' W_n \quad \text{and} \quad \mathcal{J}_0 = -\big(\dot G_0' W_0 \dot G_0\big)^{-1}\dot G_0' W_0.$$

As already mentioned, Lemma 1 effectively becomes a variant of Pakes and Pollard (1989, Theorem 3.3), when $\rho = 2$. In particular, when $\rho = 2$, condition (iv) becomes $\hat G_n(\theta_0,\hat\gamma_n) = O_{\mathbb{P}}(n^{-1/2})$, conditions (i)–(iii) and (v) reduce to natural analogs of those of Pakes and Pollard (1989, Theorem 3.3), condition (vi) becomes $\hat W_n - W_0 = o_{\mathbb{P}}(1)$ and $\dot G(\hat\gamma_n) - \dot G_0 = o_{\mathbb{P}}(1)$, and condition (vii) is implied by the other conditions of the lemma.

In Lemma 1, the magnitude of the departure from standard asymptotics is therefore governed by the parameter ρ. The introduction of this parameter is motivated by the fact that although $\hat G_n(\theta_0,\hat\gamma_n) = O_{\mathbb{P}}(n^{-1/2})$ can fail to hold under Conditions AS and AN, the weaker condition (iv) in Lemma 1 typically holds even when its $\rho = 2$ version does not.

To be more precise, when $\rho > 2$, conditions (iii) and (iv) of Lemma 1 are weaker than their $\rho = 2$ counterparts, whereas conditions (ii), (vi), and (vii) are stronger than their $\rho = 2$ counterparts. Importantly, however, the technical tools routinely applied to verify the conditions of results such as Lemma 1 in the standard (i.e., $\rho = 2$) case can also be used to verify most (if not all) of the conditions even when a failure of $\hat G_n(\theta_0,\hat\gamma_n) = O_{\mathbb{P}}(n^{-1/2})$ implies that $\rho > 2$ is required in condition (iv). In particular, even when $\rho > 2$, condition (ii) is a relatively mild smoothness condition on G and condition (iii) can be verified using standard empirical process techniques, as can the displayed part of condition (vii).

In Section 6, we illustrate how to verify the conditions of Lemma 1 with $\rho = 3$ for the case of IPW estimators with possibly non-smooth moment conditions.

REMARK 2: While the property $\dot G(\hat\gamma_n)'\hat W_n \hat G_n(\hat\theta_n,\hat\gamma_n) = o_{\mathbb{P}}(n^{-1/2})$ assumed in condition (vii) is implied by the other conditions of the lemma when $\rho = 2$, verification of this property seems to require additional conditions when $\rho > 2$. As explained in a subsection following the proof of Lemma 1, one possibility is to require that g is of dimension d_θ, while another possibility is to require $\rho < 3$ and that $o_{\mathbb{P}}(n^{-1/2})$ can be replaced by $o_{\mathbb{P}}(n^{1/\rho-1})$ in the displayed part of condition (vii).

5.2. *Condition AS*

When $\bar{g}_n$ is of the form (5), the error in the approximation

$$g_n\big(z_i, \hat{\gamma}_n^{(i)}\big) \approx \bar{g}_n\big(z_i, \hat{\gamma}_n^{(i)}\big) + g_n(z_i, \gamma_n) - \bar{g}_n(z_i, \gamma_n)$$

is usually "cubic" in $\hat{\gamma}_n^{(i)} - \gamma_n$ (in some suitable sense), in which case the first part of Condition AS is satisfied provided $\hat{\gamma}_n^{(i)} - \gamma_n = o_{\mathbb{P}}(n^{-1/6})$ (in some suitable sense). The ease with which these heuristics can be made rigorous depends in part on the smoothness of g, but suffice it to say that a condition of the form $\hat{\gamma}_n - \gamma_n = o_{\mathbb{P}}(n^{-1/6})$ has been found to be sufficient in all of the cases we have examined, including even the non-differentiable-in-γ example used in the Monte Carlo experiment of Section 7 (and analyzed in Section SA.3 of the Supplemental Material).

Whereas it is usually most efficient to proceed on a case-by-case basis when verifying the first part of Condition AS, the second part of the condition admits general sufficient conditions that are both mild and relatively simple. A common way of verifying the second part of Condition SE (i.e., the stochastic equicontinuity counterpart of Condition AS) is to exhibit a sequence Γ_n satisfying $\mathbb{P}(\hat{\gamma}_n \in \Gamma_n) \to 1$ and

$$\sup_{\gamma \in \Gamma_n} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[\bar{g}_0(z_i, \gamma) - \bar{G}_0(\gamma) - \bar{g}_0(z_i, \gamma_0) + \bar{G}_0(\gamma_0)\big] \right\| = o_{\mathbb{P}}(1),$$

where empirical process results (e.g., maximal inequalities) can be used to formulate primitive sufficient conditions for the latter (see, e.g., Andrews (1994b, Condition (3.36)), Chen, Linton, and van Keilegom (2003, Conditions (2.4) and (2.5′)), and references therein). An analogous approach does not seem applicable when the goal is to formulate primitive sufficient conditions for the second part of Condition AS, as the dependence of $\hat{\gamma}_n^{(i)}$ on i implies that the second part of Condition AS cannot be deduced with the help of a result of the form

$$\sup_{\gamma \in \Gamma_n} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[\bar{g}_n(z_i, \gamma) - \bar{G}_n(\gamma) - \bar{g}_n(z_i, \gamma_n) + \bar{G}_n(\gamma_n)\big] \right\| = o_{\mathbb{P}}(1).$$

Instead, the proof of the following lemma exploits the fact that the object of interest can be expressed as a linear combination of U-statistics when $\hat{\gamma}_n$ is kernel-based. Here, and elsewhere in the paper, it is tacitly assumed that the indices i, j, and k are distinct, unless explicitly noted otherwise.

LEMMA 2: *Suppose that $\hat{\gamma}_n$ is kernel-based, that $\bar{g}_n$ is of the form (5), and that*

$$\mathbb{V}\big(g_{n,\gamma}(z_i)\big[\hat{\gamma}_n^j - \gamma_n\big]\big) = o(n), \qquad \mathbb{V}\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^k - \gamma_n\big]\big) = o\big(n^2\big),$$
$$\mathbb{V}\big(E\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n\big]|z_i\big)\big) = o\big(n^2\big), \qquad \mathbb{V}\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j - \gamma_n, \hat{\gamma}_n^j - \gamma_n\big]\big) = o\big(n^3\big).$$

Then the second part of Condition AS is satisfied.

5.3. *Condition AN*

When $\bar{g}_n$ is of the form (5), we have

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

where

$$\psi_n(z_i) = g_n(z_i, \gamma_n) - G_n(\gamma_n) + \delta_n(z_i), \qquad \delta_n(z_i) = G_{n,\gamma}\big[\hat{\gamma}_n^i - \gamma_n\big],$$

and

$$\hat{\mathcal{B}}_n = G_n(\gamma_n) + \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} G_{n,\gamma\gamma}\big[\hat{\gamma}_n^{(i)} - \gamma_n, \hat{\gamma}_n^{(i)} - \gamma_n\big].$$

Direct calculations can usually be used to demonstrate existence of a function ψ_0 satisfying

$$\mathbb{E}\big\|\psi_n(z) - \psi_0(z)\big\|^2 \to 0, \qquad \mathbb{E}\big\|\psi_0(z)\big\|^2 < \infty. \tag{10}$$

Indeed, under general conditions, (10) holds with $\psi_0(z) = g_0(z, \gamma_0) + \delta_0(z)$, where $\delta_0(z)$ is the "correction term" discussed by Newey (1994a). If (10) holds, then Condition AN is satisfied if also $\hat{\mathcal{B}}_n = \mathcal{B}_n + o_{\mathbb{P}}(n^{-1/2})$. A simple sufficient condition for this to occur is given in the next result.

LEMMA 3: *Suppose that $\hat{\gamma}_n$ is kernel-based, that $\bar{g}_n$ is of the form (5), that (10) holds, and that*

$$\mathbb{V}\big(G_{n,\gamma\gamma}\big[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^i - \gamma_n\big]\big) = o\big(n^2\big), \qquad \mathbb{V}\big(G_{n,\gamma\gamma}\big[\hat{\gamma}_n^i - \gamma_n, \hat{\gamma}_n^j - \gamma_n\big]\big) = o(n).$$

Then Condition AN holds with $\Omega_0 = \mathbb{V}[\psi_0(z)]$ and any $\mathcal{B}_n = \mathbb{E}\hat{\mathcal{B}}_n + o(n^{-1/2})$.

5.4. *Conditions AL*, AS*, and AN**

Condition AL* can often be verified with the help of the following bootstrap analogue of Lemma 1.

LEMMA 4: *Suppose that the assumptions of Lemma 1 are satisfied, that $\hat{\theta}_n^* - \theta_0 = o_{\mathbb{P}}(1)$, and that*:

(i*) $\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) \le \inf_{\theta\in\Theta} \hat{G}_n^*(\theta, \hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\theta, \hat{\gamma}_n^*) + o_{\mathbb{P}}(n^{-1})$;

(ii*) *for every $\delta_n = o(1)$,*

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \frac{\big\|G\big(\theta, \hat{\gamma}_n^*\big) - G\big(\theta_0, \hat{\gamma}_n^*\big) - \dot{G}\big(\hat{\gamma}_n^*\big)(\theta - \theta_0)\big\|}{\|\theta - \theta_0\|^{\rho/2}} = o_{\mathbb{P}}(1);$$

(iii*) *for every $\delta_n = o(1)$,*

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \frac{\big\|\hat{G}_n^*\big(\theta, \hat{\gamma}_n^*\big) - G\big(\theta, \hat{\gamma}_n^*\big) - \hat{G}_n^*\big(\theta_0, \hat{\gamma}_n^*\big) + G\big(\theta_0, \hat{\gamma}_n^*\big)\big\|}{1 + n^{1/\rho}\|\theta - \theta_0\|} = o_{\mathbb{P}}\big(n^{-1/\rho}\big);$$

(iv*) $\hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) = O_{\mathbb{P}}(n^{-1/\rho})$;

(vi*) $\hat{W}_n^* - W_n = o_{\mathbb{P}}(n^{1/\rho-1/2})$ *and* $\dot{G}(\hat{\gamma}_n^*) - \dot{G}_n = o_{\mathbb{P}}(n^{1/\rho-1/2})$;

(vii*) $\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_{\mathbb{P}}(n^{-1/2})$ *and, for every $\delta_n = O(n^{-1/\rho})$,*

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \big\|\hat{G}_n^*\big(\theta, \hat{\gamma}_n^*\big) - G\big(\theta, \hat{\gamma}_n^*\big) - \hat{G}_n^*\big(\theta_0, \hat{\gamma}_n^*\big) + G\big(\theta_0, \hat{\gamma}_n^*\big)\big\| = o_{\mathbb{P}}\big(n^{-1/2}\big).$$

Then Condition AL holds with $\mathcal{J}_n^* = \mathcal{J}_n$ and $\mathcal{J}_0^* = \mathcal{J}_0$.*

When the first part of Condition AS is satisfied with $\bar{g}_n$ of the form (5), there usually exist linear and bilinear functionals $g^*_{n,\gamma}(z)[\cdot]$ and $g^*_{n,\gamma\gamma}(z)[\cdot,\cdot]$ such that the first part of Condition AS* is satisfied with

$$\bar{g}^*_n(z,\gamma) = g^*_n(z,\hat{\gamma}_n) + g^*_{n,\gamma}(z)[\gamma-\hat{\gamma}_n] + \frac{1}{2}g^*_{n,\gamma\gamma}(z)[\gamma-\hat{\gamma}_n,\gamma-\hat{\gamma}_n]. \tag{11}$$

Conditions under which the second part of Condition AS* holds when $\bar{g}^*_n$ is of the form (11) are given in the following bootstrap analogue of Lemma 2.

LEMMA 5: *Suppose that $\hat{\gamma}^*_n$ is kernel-based, that $\bar{g}^*_n$ is of the form* (11)*, and that*

$$\begin{aligned}
\mathbb{V}^*\big(g^*_{n,\gamma}\big(z^*_{i,n}\big)\big[\hat{\gamma}^{*,j}_n-\hat{\gamma}_n\big]\big) &= o_{\mathbb{P}}(n),\\
\mathbb{V}^*\big(g^*_{n,\gamma\gamma}\big(z^*_{i,n}\big)\big[\hat{\gamma}^{*,j}_n-\hat{\gamma}_n,\hat{\gamma}^{*,k}_n-\hat{\gamma}_n\big]\big) &= o_{\mathbb{P}}\big(n^2\big),\\
\mathbb{V}^*\big(\mathbb{E}^*\big(g^*_{n,\gamma\gamma}\big(z^*_{i,n}\big)\big[\hat{\gamma}^{*,j}_n-\hat{\gamma}_n,\hat{\gamma}^{*,j}_n-\hat{\gamma}_n\big]|z^*_{i,n}\big)\big) &= o_{\mathbb{P}}\big(n^2\big),\\
\mathbb{V}^*\big(g^*_{n,\gamma\gamma}\big(z^*_{i,n}\big)\big[\hat{\gamma}^{*,j}_n-\hat{\gamma}_n,\hat{\gamma}^{*,j}_n-\hat{\gamma}_n\big]\big) &= o_{\mathbb{P}}\big(n^3\big),
\end{aligned}$$

where $\mathbb{V}^*[\cdot]$ denotes $\mathbb{V}[\cdot|z_1,\dots,z_n]$. Then the second part of Condition AS is satisfied.*

Finally, when $\bar{g}^*_n$ is of the form (11), we have

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

where

$$\psi^*_n\big(z^*_{i,n}\big) = g^*_n\big(z^*_{i,n},\hat{\gamma}_n\big)-G^*_n(\hat{\gamma}_n)+\delta^*_n\big(z^*_{i,n}\big), \qquad \delta^*_n\big(z^*_{i,n}\big) = G^*_{n,\gamma}\big[\hat{\gamma}^{*,i}_n-\hat{\gamma}_n\big],$$

and

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

with

$$\begin{aligned}
G^*_n(\gamma) &= \mathbb{E}^*g^*_n\big(z^*_{i,n},\gamma\big), \qquad G^*_{n,\gamma}[\eta] = \mathbb{E}^*g^*_{n,\gamma}\big(z^*_{i,n}\big)[\eta],\\
G^*_{n,\gamma\gamma}[\eta,\varphi] &= \mathbb{E}^*g^*_{n,\gamma\gamma}\big(z^*_{i,n}\big)[\eta,\varphi].
\end{aligned}$$

Direct calculations can usually be used to show that

$$\mathbb{E}^*\big\|\psi^*_n\big(z^*_{i,n}\big)-\psi_n\big(z^*_{i,n}\big)\big\|^2 = o_{\mathbb{P}}(1), \tag{12}$$

in which case the following bootstrap analogue of Lemma 3 can be used to verify Condition AN*.

LEMMA 6: *Suppose that the assumptions of Lemma 3 are satisfied, that $\hat{\gamma}^*_n$ is kernel-based, that $\bar{g}^*_n$ is of the form* (11)*, that* (12) *holds, and that*

$$\mathbb{V}^*\big(G^*_{n,\gamma\gamma}\big[\hat{\gamma}^{*,i}_n-\hat{\gamma}_n,\hat{\gamma}^{*,i}_n-\hat{\gamma}_n\big]\big) = o_{\mathbb{P}}\big(n^2\big),$$

$$\mathbb{V}^*\big(G^*_{n,\gamma\gamma}\big[\hat{\gamma}^{*,i}_n - \hat{\gamma}_n, \hat{\gamma}^{*,j}_n - \hat{\gamma}_n\big]\big) = o_{\mathbb{P}}(n),$$
$$\mathbb{E}^*\hat{\mathcal{B}}^*_n = \mathbb{E}\hat{\mathcal{B}}^*_n + o_{\mathbb{P}}\big(n^{-1/2}\big).$$

Then Condition AN holds with* $\Omega^*_0 = \Omega_0$ *and any* $\mathcal{B}^*_n = \mathbb{E}\hat{\mathcal{B}}^*_n + o(n^{-1/2})$.

REMARK 3: If the conditions of Lemma 6 are satisfied, then $\hat{\Omega}_n = n^{-1}\sum_{i=1}^n \psi^*_n(z_i)\psi^*_n(z_i)'$ is a consistent estimator of Ω_0. Although $\hat{\Omega}_n$ emerges here as a by-product of our analysis of the bootstrap, it is interesting to note that it can be interpreted as a variant of the "delta-method" variance estimator of Newey (1994b).

6. EXAMPLE: INVERSE PROBABILITY WEIGHTING

In the previous sections, the average density example was chosen for illustrative purposes because it highlights exactly those parts of our high-level assumptions that differ from conventional ones, namely, Condition AN (which quantifies the departure from conventional conclusions) and the second part of Condition AS (which enables us to depart from conventional assumptions). Indeed, the estimators discussed in connection with Example 1 were intentionally chosen in such a way that Condition AL and the first part of Condition AS are representations in the sense that they hold without any $o_{\mathbb{P}}(n^{-1/2})$ and $o_{\mathbb{P}}(1)$ terms.

To substantiate the claim that Example 1 is nevertheless representative, this section examines a more substantive and complicated class of estimators, namely, IPW estimators. For these estimators, Condition AL and the first part of Condition AS are not merely representations, but as discussed in what follows, they nevertheless remain verifiable under assumptions that are sufficiently weak to permit us to obtain distributional results that differ from conventional ones, a difference that once again is quantified by Condition AN and can be brought to light thanks to the second part of Conditions AS.

Suppose $z_1, \ldots, z_n$ are i.i.d. copies of $z = (y, t, x')'$, where $y \in \mathbb{R}$ is a scalar dependent variable, $t \in \{0, 1\}$ is a binary indicator, and $x \in \mathbb{X} \subseteq \mathbb{R}^d$ is a continuous covariate with density f_0. Assuming the estimand $\theta_0 \in \Theta \subseteq \mathbb{R}^{d_\theta}$ is the unique solution to an equation of the form

$$\mathbb{E}\left[\frac{t}{q_0(x)} m(y; \theta)\right] = 0, \quad q_0(x) = \mathbb{E}(t|x) = \mathbb{P}[t = 1|x],$$

where m is a known $\mathbb{R}^{d_\theta}$-valued function, an IPW estimator $\hat{\theta}_n$ of θ_0 is one that satisfies

$$\frac{1}{n}\sum_{i=1}^n \frac{t_i}{\hat{q}_n(x_i)} m(y_i; \hat{\theta}_n) = o_{\mathbb{P}}\big(n^{-1/2}\big),$$

where $\hat{q}_n$ is an estimator of (the propensity score) q_0.

In what follows, we assume that q_0 is estimated using a local polynomial estimator of order $P > 3d/4 - 1$. To describe this estimator, define $d_P = (P + d - 1)!/[P!(d - 1)!]$, and let $b_P(x) \in \mathbb{R}^{d_P}$ denote the Pth-order polynomial basis expansion based on $x =$

$(x_1, \ldots, x_d)' \in \mathbb{R}^d$; that is,

$$b_P(x) = \begin{pmatrix} 1 \\ [x]^1 \\ \vdots \\ [x]^P \end{pmatrix}, \quad [x]^p = \begin{pmatrix} x_1^p \\ x_1^{p-1}x_2 \\ \vdots \\ x_d^p \end{pmatrix}.$$

Also, let

$$\hat{\gamma}_{x,n}(x) = \operatorname{vec}_P\left[\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{x,n}(x_i - x)\right], \qquad \mathcal{K}_{x,n}(u) = b_{P,n}(u)b_{P,n}(u)'K_n(u),$$

and

$$\hat{\gamma}_{t,n}(x) = \frac{1}{n}\sum_{i=1}^n t_i \mathcal{K}_{t,n}(x_i - x), \quad \mathcal{K}_{t,n}(u) = b_{P,n}(u)K_n(u),$$

where $b_{P,n}(u) = b_P(u/h_n)$, $K_n(u) = K(u/h_n)/h_n^d$, h_n is a bandwidth, K is a kernel, and where $\operatorname{vec}_P : \mathbb{R}^{d_P \times d_P} \to \mathbb{R}^{d_P^2}$ is the vectorization operator. The Pth-order local polynomial estimator of $q_0(x)$ is given by $q(x; \hat{\gamma}_n)$, where

$$q(x;\gamma) = e_P'\big(\operatorname{vec}_P^{-1}\big[\gamma_x(x)\big]\big)^{-1}\gamma_t(x), \quad \gamma = \big(\gamma_x', \gamma_t'\big)',$$

e_P is the first unit vector in $\mathbb{R}^{d_P}$, and $\operatorname{vec}_P^{-1} : \mathbb{R}^{d_P^2} \to \mathbb{R}^{d_P \times d_P}$ is the inverse of vec_P.

Because $\hat{\gamma}_n$ is kernel-based, the associated IPW estimator $\hat{\theta}_n$ is a kernel-based two-step semiparametric, which can be analyzed using the results of the previous sections by representing the defining property of $\hat{\theta}_n$ as

$$\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}\big(n^{-1}\big), \quad \hat{W}_n = I_{d_\theta},$$

where

$$g(z, \theta, \gamma) = \frac{t}{q(x;\gamma)}m(y;\theta)$$

is neither linear in γ nor (necessarily) differentiable in θ. Doing so, it is shown in Section A.2 of the Supplemental Material that under regularity conditions and if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and $nh_n^{2P+2} \to 0$, then the conditions of Theorems 1 and 2 are satisfied. In what follows, we briefly describe the main steps in the proof(s).

First, consider Condition AL. Under the stated bandwidth conditions, it follows from the discussion below that $\hat{G}_n(\theta_0, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/3})$. Accordingly, we set $\rho = 3$ when verifying Condition AL with the help of Lemma 1. To define the other main objects of that lemma, set $W_n = W_0 = I_{d_\theta}$ and let

$$\gamma_{x,n}(x) = \operatorname{vec}_P\left[\int_{\mathbb{R}^d} \mathcal{K}_x(u)f_0(x + uh_n)\,du\right], \quad \mathcal{K}_x(u) = b_P(u)b_P(u)'K(u),$$

$$\gamma_{t,n}(x) = \int_{\mathbb{R}^d} \mathcal{K}_t(u)q_0(x + uh_n)f_0(x + uh_n)\,du, \quad \mathcal{K}_t(u) = b_P(u)K(u),$$

and

$$\gamma_{x,0}(x) = f_0(x)\operatorname{vec}_P\left[\int_{\mathbb{R}^d} \mathcal{K}_x(u)\,du\right], \qquad \gamma_{t,0}(x) = q_0(x)f_0(x)\int_{\mathbb{R}^d} \mathcal{K}_t(u)\,du.$$

The functional G can be represented as

$$G(\theta,\gamma) = \mathbb{E}\left[\frac{q_0(x)}{q(x;\gamma)} r_0(x;\theta)\right], \quad r_0(x;\theta) = \mathbb{E}\big[m(y;\theta)|x, t=1\big],$$

and satisfies $G(\theta,\gamma_0) = 0$ if and only if $\theta = \theta_0$ because $q(x;\gamma_0) = q_0(x)$. Moreover, under regularity conditions, including differentiability of $r_0(x;\cdot)$, we have

$$\dot{G}(\gamma) = \mathbb{E}\left[\frac{q_0(x)}{q(x;\gamma)}\dot{r}_0(x)\right], \quad \dot{r}_0(x) = \frac{\partial}{\partial\theta} r_0(x;\theta)\Big|_{\theta=\theta_0}.$$

Apart from condition (iv), the hardest-to-verify conditions of Lemma 1 are (iii) and the displayed part of (vii). We verify these conditions with the help of empirical process techniques and using the fact that

$$\max_{1\le i\le n}\big\|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\big\| = o_{\mathbb{P}}\big(n^{-1/6}\big)$$

when $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$.

Next, consider Condition AS. Because $g(z,\theta,\gamma)$ is a smooth functional of γ, it is natural to set $\bar{g}_n$ equal to a second-order Taylor approximation to g_n obtained by expanding around $\gamma = \gamma_n$. Simple bounding arguments can be used to show that the resulting $\bar{g}_n$ satisfies the first part of Condition AS because $\max_{1\le i\le n}\|\hat{\gamma}_n(x_i) - \gamma_n(x_i)\| = o_{\mathbb{P}}(n^{-1/6})$. Moreover, because $\bar{g}_n$ is of the form (5), Lemma 2 can be used to show that the second part of Condition AS is satisfied whenever $nh_n^d \to \infty$.

Condition AN is also satisfied, as can be shown using Lemma 3. To be specific, (10) holds with

$$\psi_0(z) = \frac{t}{q_0(x)}m(y;\theta_0) - \frac{r_0(x;\theta_0)}{q_0(x)}\big(t - q_0(x)\big),$$

while lengthy calculations show that if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and $nh_n^{2P+2} \to 0$, then we can set

$$\begin{aligned}\mathcal{B}_n = {} & -\frac{K(0)}{nh_n^d}\big(e_P'\Gamma_x^{-1}e_P\big)\int_{\mathbb{X}}\frac{1-q_0(u)}{q_0(u)}r_0(u;\theta_0)\,du \\ & + \frac{1}{nh_n^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{r_0(u;\theta_0)f_0(u)}{q_0(u)^2}e_P' \\ & \times \Gamma_{x,n}(u)^{-1}\mathcal{K}_t(v)\mathcal{K}_t(v)'\Gamma_{x,n}(u)^{-1}e_P\sigma_t^2(u+vh_n)f_0(u+vh_n)\,du\,dv,\end{aligned}$$

where

$$\Gamma_{x,n}(x) = \operatorname{vec}_P^{-1}\big(\gamma_{x,n}(x)\big), \qquad \Gamma_x = \int_{\mathbb{R}^d}\mathcal{K}_x(u)\,du, \qquad \sigma_t^2(x) = q_0(x)\big(1-q_0(x)\big).$$

Because Conditions AN and AL both hold, with $\|\mathcal{B}_n\| = O(n^{-1}h_n^{-d})$ in the latter, we have $\hat{G}_n(\theta_0,\hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/2} + \|\mathcal{B}_n\|) = O_{\mathbb{P}}(n^{-1/3})$ when $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$. In other words, condition (iv) of Lemma 1 holds with $\rho = 3$.

To summarize, if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and if $nh_n^{2P+2} \to 0$, then the conditions of Theorem 1 are satisfied and (2) holds with

$$\mathfrak{B}_n = -\dot{G}_n^{-1}\mathcal{B}_n, \quad \dot{G}_n = \mathbb{E}\left[\frac{q_0(x)}{q(x;\gamma_n)}\dot{r}_0(x)\right],$$

and

$$\Sigma_0 = \dot{G}_0^{-1}\mathbb{V}\big[\psi_0(z)\big]\dot{G}_0^{-1}, \quad \dot{G}_0 = \mathbb{E}\big[\dot{r}_0(x)\big].$$

Proceeding in a similar way, Conditions AL*, AS*, and AN* can be verified using Lemmas 4, 5, and 6, respectively. Moreover, $\mathcal{B}_n^*$ can be set equal to $\mathcal{B}_n$ in Lemma 6, so it follows from Theorem 2 that the bootstrap consistency condition (7) is satisfied.

Importantly, while perhaps not the weakest possible, the bandwidth conditions we impose are sufficiently weak to permit $\hat{\theta}_n$ to exhibit a nonnegligible asymptotic bias. To be specific, the bandwidth condition $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ allows for the possibility that $nh_n^{2d} \nrightarrow \infty$, in which case $\mathfrak{B}_n = O(n^{-1}h_n^{-d}) \neq o(n^{-1/2})$.

7. SIMULATION EVIDENCE

We conducted a small-scale Monte Carlo experiment to explore some of the implications of our theoretical results in samples of moderate size. Because the simulation study involves bootstrap procedures, computational considerations let us to consider a closed form estimator and a relatively small sample size.

The estimator we consider is the one previously analyzed in the Hit Rate example of Chen, Linton, and van Keilegom (2003), which we also re-analyze using our results in Section SA.3 of the Supplemental Material. To describe this estimator, let $z_1, \ldots, z_n$ be i.i.d. copies of $z = (y, x')'$, where $y \in \mathbb{R}$ is a scalar dependent variable and $x \in \mathbb{R}^d$ is a continuous covariate with density γ_0. The parameter of interest is the scalar $\theta_0 = \mathbb{P}[y \geq \gamma_0(x)] = \mathbb{E}[\mathbb{1}(y \geq \gamma_0(x))]$, a kernel-based semiparametric estimator of which is given by

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\big(y_i \geq \hat{\gamma}_n(x_i)\big), \quad \hat{\gamma}_n(x) = \frac{1}{n}\sum_{j=1}^{n}K_n(x_j - x),$$

where $K_n(x) = K(x/h_n)/h_n^d$, h_n is a bandwidth, and K is a kernel.

Although the estimator $\hat{\theta}_n$ is in closed form (i.e., satisfies Condition AL without any $o_{\mathbb{P}}(n^{-1/2})$ term), the estimator is significantly more complicated than the average density estimators of Example 1 because it is a non-smooth functional of $\hat{\gamma}_n$. Nevertheless, it is shown in Section SA.3 of the Supplemental Material that $\hat{\theta}_n$ can be analyzed using the results of this paper. In particular, under the regularity conditions given there, we show that if $nh_n^{3d/2}/(\log n)^{3/2} \to \infty$ and if $nh_n^{2P} \to 0$, with P the kernel order, then the conditions of Theorems 1 and 2 are satisfied with $\Sigma_0^* = \Sigma_0$ and $\mathfrak{B}_n^* = \mathfrak{B}_n = O[1/(nh_n^d)]$. The explicit formulas for all the biases and variance quantities are given in the Supplemental Material for brevity.

We consider $S = 1{,}000$ replications for the Monte Carlo experiment, where for each replication we generate a random sample of size $n = 1{,}000$ from a model of the form

$$\begin{pmatrix} y_i \\ x_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_y \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & 0' \\ 0 & \sigma_x^2 I_d \end{pmatrix}\right).$$

TABLE I

SIMULATION DATA GENERATING PROCESSES

Model	d	P	μ_y	σ_y	σ_x	θ_0	Σ_0	$\mathcal{B}_0^{\mathrm{SB}}$	$\mathcal{B}_0^{\mathrm{S}}$	$h_{\mathtt{opt}}$
M1	1	2	1/2	1/2	1/2	0.449	0.322	−0.253	0.402	0.086
M2	1	4	1/2	1/2	1/2	0.449	0.322	−0.374	0.544	0.233
M3	2	2	1/2	1/2	1/2	0.633	0.296	−0.115	0.517	0.122
M4	2	4	1/2	1/2	1/2	0.633	0.296	−0.260	0.826	0.261
M5	3	4	1/2	1/2	1/2	0.732	0.226	−0.142	0.681	0.298
M6	1	2	1/2	1/3	1/3	0.229	0.209	−0.079	0.101	0.092
M7	1	4	1/2	1/3	1/3	0.229	0.209	−0.092	−4.072	0.089
M8	2	2	1/2	1/3	1/3	0.356	0.305	−0.079	0.585	0.108
M9	2	4	1/2	1/3	1/3	0.356	0.305	−0.162	−3.960	0.165
M10	3	4	1/2	1/3	1/3	0.457	0.358	−0.127	−2.195	0.238
M11	1	2	1/2	1/4	1/3	0.208	0.203	−0.070	−0.297	0.049
M12	1	4	1/2	1/4	1/3	0.208	0.203	−0.077	−7.057	0.077
M13	2	2	1/2	1/4	1/3	0.351	0.319	−0.081	0.278	0.131
M14	2	4	1/2	1/4	1/3	0.351	0.319	−0.167	−6.784	0.152
M15	3	4	1/2	1/4	1/3	0.464	0.385	−0.133	−4.332	0.218
M16	1	2	3/4	3/4	1/4	0.327	0.283	−0.108	1.318	0.043
M17	1	4	3/4	3/4	1/4	0.327	0.283	−0.153	3.338	0.136
M18	2	2	3/4	3/4	1/4	0.318	0.266	−0.040	1.018	0.079
M19	2	4	3/4	3/4	1/4	0.318	0.266	−0.084	−7.957	0.132
M20	3	4	3/4	3/4	1/4	0.328	0.261	−0.053	−15.813	0.158
M21	1	2	1	1/2	1/5	0.280	0.276	−0.049	1.029	0.036
M22	1	4	1	1/2	1/5	0.280	0.276	−0.054	−27.472	0.055
M23	2	2	1	1/2	1/5	0.252	0.229	−0.029	−1.514	0.066
M24	2	4	1	1/2	1/5	0.252	0.229	−0.060	−73.575	0.086
M25	3	4	1	1/2	1/5	0.241	0.210	−0.035	−74.966	0.120

As described in Table I, a total of 25 different configurations of μ_y, σ_y^2, σ_x^2, d, and P were considered. Some of these models (namely, those with $(d, P) \in \{(1, 2), (2, 2)\}$) are not covered by conventional first-order asymptotic results (because P is too small), but because our large sample results only require $P > 3d/4$, all of the models listed in Table I are covered by the results of this paper.

We focus on the performance of three 95% confidence intervals, namely, the (feasible) bootstrap-based intervals $\mathsf{CI}_{0.95}^{\mathrm{E}}$ and $\mathsf{CI}_{0.95}^{\mathrm{P}}$ and an infeasible version of $\mathsf{CI}_{0.95}^{\mathrm{N}}$ obtained by setting $\hat{\Sigma}_n$ equal to n times the simulation variance of $\hat{\theta}_n$. We use the simulation variance of $\hat{\theta}_n$ to avoid rendering our results sensitive to the choice of additional tuning parameters needed in order to estimate the (complicated) asymptotic variance of $\hat{\theta}_n$. In the simulations, for each replication we approximate the bootstrap distribution by resampling $B = 1{,}000$ times. For each model, we report results for a range of bandwidths h_n, partly with the aim of judging the relevance of one of the main predictions of our theory (e.g., Proposition 1), namely, that

$$\mathbb{P}\big[\theta_0 \in \mathsf{CI}_{0.95}^{\mathrm{E}}\big] \leq \mathbb{P}\big[\theta_0 \in \mathsf{CI}_{0.95}^{\mathrm{N}}\big] \leq \mathbb{P}\big[\theta_0 \in \mathsf{CI}_{0.95}^{\mathrm{P}}\big] \approx 0.95,$$

with strict inequalities for "small" bandwidths, that is, whenever $nh_n^{2d} \nrightarrow \infty$.

Tables II–VI report the main results. For each model, we consider a grid of bandwidths of the form $h_n = c \cdot h_{\mathtt{opt}}$, where $c \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1\}$ and where $h_{\mathtt{opt}}$ is an "optimal" (in a certain sense) bandwidth characterized in Section 3.4 of the

TABLE II

SIMULATION RESULTS FOR MODELS M1–M5 ($n = 1{,}000$, $B = 1{,}000$, $S = 1{,}000$)[a]

	BW		CR			IL				
	h_n	c	E	N	P	E	N	P	B/SE	MSE
M1: $d = 1, P = 2$	0.043	0.5	0.930	0.942	0.935	0.072	0.070	0.072	−0.258	1.070
	0.051	0.6	0.942	0.944	0.939	0.072	0.070	0.072	−0.185	1.034
	0.060	0.7	0.947	0.948	0.942	0.072	0.070	0.072	−0.124	1.014
	0.069	0.8	0.952	0.948	0.939	0.072	0.070	0.072	−0.067	1.003
	0.077	0.9	0.954	0.946	0.940	0.071	0.070	0.071	−0.016	1.000
$h_{\mathtt{opt}} =$	0.086	1.0	0.955	0.946	0.939	0.071	0.070	0.071	0.032	1.000
	0.094	1.1	0.957	0.946	0.937	0.071	0.070	0.071	0.084	1.003
M2: $d = 1, P = 4$	0.117	0.5	0.949	0.944	0.940	0.072	0.070	0.072	−0.141	1.015
	0.140	0.6	0.952	0.946	0.941	0.072	0.070	0.072	−0.104	1.007
	0.163	0.7	0.951	0.947	0.941	0.071	0.070	0.071	−0.070	1.001
	0.186	0.8	0.954	0.949	0.940	0.071	0.070	0.071	−0.041	0.999
	0.210	0.9	0.954	0.949	0.940	0.071	0.070	0.071	−0.010	0.996
$h_{\mathtt{opt}} =$	0.233	1.0	0.953	0.948	0.941	0.071	0.070	0.071	0.025	1.000
	0.256	1.1	0.953	0.946	0.939	0.071	0.070	0.071	0.065	1.004
M3: $d = 2, P = 2$	0.061	0.5	0.082	0.657	0.935	0.073	0.068	0.073	−1.606	3.709
	0.073	0.6	0.407	0.828	0.936	0.072	0.068	0.072	−1.022	2.114
	0.085	0.7	0.688	0.899	0.932	0.071	0.068	0.071	−0.644	1.457
	0.098	0.8	0.833	0.933	0.929	0.070	0.068	0.070	−0.373	1.158
	0.110	0.9	0.904	0.946	0.927	0.069	0.067	0.069	−0.155	1.030
$h_{\mathtt{opt}} =$	0.122	1.0	0.943	0.948	0.918	0.069	0.067	0.069	0.033	1.000
	0.134	1.1	0.959	0.944	0.903	0.068	0.067	0.068	0.201	1.033
M4: $d = 2, P = 4$	0.130	0.5	0.642	0.873	0.939	0.071	0.069	0.071	−0.810	1.724
	0.156	0.6	0.814	0.921	0.943	0.071	0.068	0.071	−0.534	1.328
	0.183	0.7	0.887	0.934	0.941	0.070	0.068	0.070	−0.353	1.152
	0.209	0.8	0.926	0.945	0.943	0.069	0.068	0.069	−0.219	1.065
	0.235	0.9	0.942	0.952	0.941	0.069	0.068	0.069	−0.103	1.024
$h_{\mathtt{opt}} =$	0.261	1.0	0.953	0.953	0.937	0.068	0.067	0.068	0.005	1.000
	0.287	1.1	0.958	0.951	0.931	0.068	0.067	0.068	0.114	1.006
M5: $d = 3, P = 4$	0.149	0.5	0.000	0.242	0.931	0.070	0.063	0.070	−2.669	9.069
	0.179	0.6	0.123	0.682	0.936	0.066	0.062	0.066	−1.511	3.544
	0.209	0.7	0.543	0.861	0.941	0.064	0.061	0.064	−0.901	1.903
	0.238	0.8	0.799	0.910	0.939	0.062	0.060	0.062	−0.528	1.317
	0.268	0.9	0.897	0.939	0.935	0.061	0.060	0.061	−0.266	1.086
$h_{\mathtt{opt}} =$	0.298	1.0	0.935	0.948	0.930	0.060	0.060	0.060	−0.060	1.000
	0.328	1.1	0.947	0.948	0.918	0.059	0.059	0.059	0.125	1.005

[a](i) Columns under BW report grid of bandwidths and multiplicative factor c relative to $h_{\mathtt{opt}}$; (ii) columns under CR report coverage error for 95% confidence intervals; (iii) columns under IL report average interval length for 95% confidence intervals; and (iv) columns B/SE and MSE report, respectively, simulation bias relative to simulation standard error and simulation mean squared error of $\hat{\theta}_n(h_n)$.

Supplemental Material. For implementation, we set $K(u) = k(u_1)k(u_2)\cdots k(u_d)$ for $u = (u_1, u_2, \ldots, u_d)' \in \mathbb{R}^d$, with $k(\cdot)$ a Pth-order univariate kernel, where $k(v) = \phi(v)$ if $P = 2$ and $k(v) = (3 - v^2)\phi(v)/2$ if $P = 4$, and $\phi(v) = d\Phi(v)/dv$. Each table includes coverage rates and average interval length for three 95% confidence intervals $\mathsf{CI}^{\mathrm{E}}_{0.95}$, $\mathsf{CI}^{\mathrm{N}}_{0.95}$, and $\mathsf{CI}^{\mathrm{P}}_{0.95}$, as well as the bias divided by the square root of the simulation variance (B/SE) and the mean squared error (MSE) of each estimator $\hat{\theta}_n$. The simulations are time consuming because for each bandwidth and each simulation replication, we need to approximate the standard (bootstrap) distribution of $\hat{\theta}^*_n$. For this reason, we focus exclusively

TABLE III

SIMULATION RESULTS FOR MODELS M6–M10 ($n = 1{,}000$, $B = 1{,}000$, $S = 1{,}000$)[a]

	BW		CR			IL				
	h_n	c	E	N	P	E	N	P	B/SE	MSE
M6: $d = 1, P = 2$	0.046	0.5	0.956	0.943	0.933	0.058	0.056	0.058	−0.095	1.004
	0.055	0.6	0.956	0.945	0.937	0.058	0.056	0.058	−0.066	1.000
	0.064	0.7	0.958	0.944	0.938	0.057	0.056	0.057	−0.040	0.999
	0.074	0.8	0.958	0.945	0.938	0.057	0.056	0.057	−0.014	0.998
	0.083	0.9	0.957	0.947	0.942	0.057	0.056	0.057	0.016	0.997
$h_{\mathtt{opt}} =$	0.092	1.0	0.958	0.946	0.943	0.057	0.056	0.057	0.048	1.000
	0.101	1.1	0.958	0.944	0.941	0.057	0.056	0.057	0.084	1.009
M7: $d = 1, P = 4$	0.045	0.5	0.957	0.944	0.927	0.058	0.056	0.058	−0.143	1.001
	0.053	0.6	0.959	0.944	0.933	0.058	0.056	0.058	−0.118	0.999
	0.062	0.7	0.957	0.944	0.930	0.058	0.056	0.058	−0.102	1.005
	0.071	0.8	0.958	0.944	0.934	0.058	0.056	0.058	−0.092	1.005
	0.080	0.9	0.958	0.945	0.935	0.058	0.056	0.058	−0.088	1.000
$h_{\mathtt{opt}} =$	0.089	1.0	0.959	0.945	0.937	0.058	0.056	0.058	−0.084	1.000
	0.098	1.1	0.957	0.945	0.937	0.058	0.056	0.058	−0.084	0.997
M8: $d = 2, P = 2$	0.054	0.5	0.116	0.674	0.896	0.065	0.066	0.065	−1.504	2.931
	0.065	0.6	0.477	0.839	0.913	0.067	0.067	0.067	−0.952	1.761
	0.075	0.7	0.730	0.915	0.913	0.068	0.067	0.068	−0.584	1.266
	0.086	0.8	0.863	0.941	0.916	0.069	0.068	0.069	−0.315	1.054
	0.097	0.9	0.927	0.950	0.915	0.069	0.068	0.069	−0.091	0.978
$h_{\mathtt{opt}} =$	0.108	1.0	0.953	0.951	0.903	0.070	0.069	0.070	0.113	1.000
	0.118	1.1	0.960	0.940	0.890	0.070	0.070	0.070	0.308	1.100
M9: $d = 2, P = 4$	0.083	0.5	0.193	0.709	0.899	0.066	0.066	0.066	−1.408	2.433
	0.099	0.6	0.533	0.830	0.912	0.067	0.067	0.067	−0.977	1.633
	0.116	0.7	0.725	0.884	0.922	0.068	0.067	0.068	−0.731	1.294
	0.132	0.8	0.817	0.910	0.926	0.068	0.067	0.068	−0.580	1.128
	0.149	0.9	0.867	0.926	0.928	0.068	0.067	0.068	−0.479	1.047
$h_{\mathtt{opt}} =$	0.165	1.0	0.898	0.930	0.932	0.068	0.067	0.068	−0.417	1.000
	0.182	1.1	0.914	0.941	0.934	0.068	0.067	0.068	−0.373	0.972
M10: $d = 3, P = 4$	0.119	0.5	0.000	0.008	0.832	0.065	0.068	0.065	−4.375	14.886
	0.143	0.6	0.001	0.312	0.887	0.070	0.071	0.070	−2.441	5.643
	0.167	0.7	0.138	0.675	0.913	0.073	0.072	0.073	−1.514	2.760
	0.190	0.8	0.502	0.831	0.923	0.074	0.073	0.074	−0.989	1.696
	0.214	0.9	0.753	0.909	0.927	0.075	0.074	0.075	−0.626	1.226
$h_{\mathtt{opt}} =$	0.238	1.0	0.883	0.935	0.933	0.076	0.075	0.076	−0.324	1.000
	0.262	1.1	0.936	0.950	0.924	0.077	0.076	0.077	−0.029	0.938

[a](i) Columns under BW report grid of bandwidths and multiplicative factor c relative to $h_{\mathtt{opt}}$; (ii) columns under CR report coverage error for 95% confidence intervals; (iii) columns under IL report average interval length for 95% confidence intervals; and (iv) columns B/SE and MSE report, respectively, simulation bias relative to simulation standard error and simulation mean squared error of $\hat{\theta}_n(h_n)$.

on a few low-dimension models, $d \in \{1, 2, 3\}$, although we did experiment with higher dimensions and found that the results reported herein are exacerbated as the dimension increases, which is not surprising (given the structure of the "small" bandwidth bias) but nevertheless important from a practical point of view.

Overall, the bootstrap-based confidence interval $\mathsf{CI}^{\mathrm{P}}_{0.95}$ performs better than its rivals in the simulations. In particular, and as predicted by our theory, the automatic bias reduction property of $\mathsf{CI}^{\mathrm{P}}_{0.95}$ established in this paper for "small" bandwidths is found to

TABLE IV

SIMULATION RESULTS FOR MODELS M11–M15 ($n = 1{,}000$, $B = 1{,}000$, $S = 1{,}000$)[a]

	BW		CR			IL				
	h_n	c	E	N	P	E	N	P	B/SE	MSE
M11: $d = 1, P = 2$	0.024	0.5	0.950	0.940	0.929	0.056	0.054	0.056	−0.224	1.033
	0.029	0.6	0.956	0.943	0.928	0.056	0.054	0.056	−0.194	1.025
	0.034	0.7	0.959	0.944	0.933	0.056	0.054	0.056	−0.176	1.014
	0.039	0.8	0.959	0.946	0.931	0.056	0.054	0.056	−0.164	1.009
	0.044	0.9	0.960	0.947	0.934	0.056	0.054	0.056	−0.156	1.007
$h_{\mathtt{opt}} =$	0.049	1.0	0.959	0.946	0.934	0.056	0.053	0.056	−0.153	1.000
	0.054	1.1	0.960	0.946	0.932	0.055	0.053	0.055	−0.151	0.995
M12: $d = 1, P = 4$	0.039	0.5	0.964	0.943	0.931	0.057	0.054	0.057	−0.151	0.995
	0.046	0.6	0.965	0.947	0.931	0.057	0.054	0.057	−0.127	0.997
	0.054	0.7	0.966	0.953	0.932	0.057	0.054	0.057	−0.110	0.999
	0.062	0.8	0.967	0.952	0.929	0.057	0.054	0.057	−0.101	1.002
	0.069	0.9	0.965	0.952	0.933	0.057	0.054	0.057	−0.096	1.003
$h_{\mathtt{opt}} =$	0.077	1.0	0.965	0.952	0.934	0.056	0.054	0.056	−0.093	1.000
	0.085	1.1	0.965	0.948	0.933	0.056	0.054	0.056	−0.093	0.992
M13: $d = 2, P = 2$	0.065	0.5	0.424	0.817	0.905	0.067	0.066	0.067	−1.052	1.845
	0.078	0.6	0.719	0.898	0.916	0.068	0.067	0.068	−0.657	1.278
	0.092	0.7	0.862	0.941	0.923	0.068	0.067	0.068	−0.382	1.025
	0.105	0.8	0.923	0.950	0.924	0.069	0.067	0.069	−0.157	0.927
	0.118	0.9	0.950	0.949	0.921	0.069	0.068	0.069	0.051	0.920
$h_{\mathtt{opt}} =$	0.131	1.0	0.963	0.941	0.910	0.070	0.069	0.070	0.259	1.000
	0.144	1.1	0.955	0.922	0.884	0.070	0.070	0.070	0.480	1.182
M14: $d = 2, P = 4$	0.076	0.5	0.042	0.593	0.882	0.065	0.065	0.065	−1.750	2.937
	0.091	0.6	0.334	0.766	0.903	0.067	0.066	0.067	−1.213	1.859
	0.106	0.7	0.589	0.848	0.909	0.068	0.067	0.068	−0.913	1.395
	0.122	0.8	0.735	0.886	0.918	0.068	0.067	0.068	−0.735	1.176
	0.137	0.9	0.812	0.904	0.919	0.068	0.067	0.068	−0.624	1.062
$h_{\mathtt{opt}} =$	0.152	1.0	0.854	0.913	0.923	0.068	0.067	0.068	−0.559	1.000
	0.167	1.1	0.875	0.919	0.927	0.068	0.067	0.068	−0.527	0.965
M15: $d = 3, P = 4$	0.109	0.5	0.000	0.000	0.750	0.059	0.066	0.059	−6.304	20.100
	0.131	0.6	0.000	0.063	0.864	0.068	0.070	0.068	−3.478	7.287
	0.152	0.7	0.005	0.412	0.904	0.072	0.072	0.072	−2.172	3.351
	0.174	0.8	0.183	0.684	0.920	0.074	0.073	0.074	−1.477	1.915
	0.196	0.9	0.490	0.813	0.926	0.075	0.073	0.075	−1.055	1.297
$h_{\mathtt{opt}} =$	0.218	1.0	0.714	0.884	0.927	0.075	0.074	0.075	−0.759	1.000
	0.239	1.1	0.840	0.918	0.933	0.076	0.075	0.076	−0.518	0.821

[a](i) Columns under BW report grid of bandwidths and multiplicative factor c relative to $h_{\mathtt{opt}}$; (ii) columns under CR report coverage error for 95% confidence intervals; (iii) columns under IL report average interval length for 95% confidence intervals; and (iv) columns B/SE and MSE report, respectively, simulation bias relative to simulation standard error and simulation mean squared error of $\hat{\theta}_n(h_n)$.

be quantitatively important. Furthermore, even when the bias appears to be small, $\mathsf{CI}^{\mathrm{P}}_{0.95}$ continues to exhibit good properties.

More specifically, our findings show that for $d = 1$, all three inference procedures perform well, as the bias highlighted in this paper is of relatively small importance. On the other hand, and more importantly, for $d = 2$ we find an important bias for "small" bandwidths. This bias is accounted for when using the percentile bootstrap (i.e., $\mathsf{CI}^{\mathrm{P}}_{0.95}$), but not when using the Efron's bootstrap (i.e., $\mathsf{CI}^{\mathrm{E}}_{0.95}$) or the infeasible version of $\mathsf{CI}^{\mathrm{N}}_{0.95}$ that employs the actual simulation (unknown in practice) variance of the estimator. Indeed,

TABLE V

SIMULATION RESULTS FOR MODELS M16–M20 ($n = 1{,}000$, $B = 1{,}000$, $S = 1{,}000$)[a]

	BW		CR			IL				
	h_n	c	E	N	P	E	N	P	B/SE	MSE
M16: $d = 1, P = 2$	0.022	0.5	0.934	0.946	0.929	0.067	0.065	0.067	−0.239	1.038
	0.026	0.6	0.943	0.947	0.931	0.067	0.066	0.067	−0.174	1.018
	0.030	0.7	0.948	0.948	0.934	0.067	0.066	0.067	−0.120	1.007
	0.035	0.8	0.951	0.948	0.933	0.067	0.066	0.067	−0.069	0.999
	0.039	0.9	0.955	0.949	0.938	0.067	0.066	0.067	−0.024	0.996
$h_{\mathtt{opt}} =$	0.043	1.0	0.956	0.949	0.934	0.067	0.066	0.067	0.020	1.000
	0.048	1.1	0.957	0.949	0.938	0.067	0.066	0.067	0.065	1.003
M17: $d = 1, P = 4$	0.068	0.5	0.952	0.947	0.937	0.067	0.066	0.067	−0.113	0.994
	0.081	0.6	0.954	0.948	0.938	0.067	0.066	0.067	−0.084	0.995
	0.095	0.7	0.956	0.949	0.941	0.067	0.066	0.067	−0.058	0.994
	0.108	0.8	0.956	0.950	0.940	0.067	0.066	0.067	−0.032	0.992
	0.122	0.9	0.955	0.948	0.940	0.067	0.066	0.067	−0.004	0.994
$h_{\mathtt{opt}} =$	0.136	1.0	0.955	0.946	0.940	0.067	0.066	0.067	0.032	1.000
	0.149	1.1	0.955	0.947	0.939	0.067	0.066	0.067	0.077	1.008
M18: $d = 2, P = 2$	0.040	0.5	0.125	0.670	0.895	0.061	0.062	0.061	−1.480	2.855
	0.048	0.6	0.474	0.847	0.913	0.063	0.063	0.063	−0.937	1.723
	0.056	0.7	0.726	0.911	0.917	0.064	0.064	0.064	−0.584	1.254
	0.063	0.8	0.857	0.937	0.918	0.065	0.064	0.065	−0.320	1.055
	0.071	0.9	0.921	0.952	0.917	0.065	0.065	0.065	−0.106	0.989
$h_{\mathtt{opt}} =$	0.079	1.0	0.948	0.952	0.908	0.066	0.066	0.066	0.093	1.000
	0.087	1.1	0.959	0.941	0.902	0.066	0.066	0.066	0.279	1.077
M19: $d = 2, P = 4$	0.066	0.5	0.344	0.787	0.907	0.063	0.063	0.063	−1.193	2.079
	0.079	0.6	0.634	0.872	0.913	0.064	0.063	0.064	−0.828	1.474
	0.092	0.7	0.784	0.909	0.924	0.064	0.063	0.064	−0.615	1.218
	0.105	0.8	0.856	0.926	0.929	0.064	0.064	0.064	−0.483	1.094
	0.119	0.9	0.891	0.932	0.933	0.065	0.064	0.065	−0.390	1.037
$h_{\mathtt{opt}} =$	0.132	1.0	0.916	0.937	0.932	0.065	0.064	0.065	−0.328	1.000
	0.145	1.1	0.929	0.941	0.934	0.065	0.064	0.065	−0.274	0.975
M20: $d = 3, P = 4$	0.079	0.5	0.000	0.000	0.347	0.045	0.053	0.045	−7.498	19.190
	0.095	0.6	0.000	0.012	0.796	0.054	0.057	0.054	−4.172	7.176
	0.111	0.7	0.000	0.260	0.868	0.059	0.060	0.059	−2.618	3.329
	0.127	0.8	0.062	0.543	0.888	0.061	0.061	0.061	−1.813	1.902
	0.143	0.9	0.300	0.729	0.904	0.062	0.062	0.062	−1.362	1.293
$h_{\mathtt{opt}} =$	0.158	1.0	0.542	0.809	0.915	0.063	0.062	0.063	−1.085	1.000
	0.174	1.1	0.684	0.853	0.919	0.064	0.063	0.064	−0.902	0.848

[a](i) Columns under BW report grid of bandwidths and multiplicative factor c relative to $h_{\mathtt{opt}}$; (ii) columns under CR report coverage error for 95% confidence intervals; (iii) columns under IL report average interval length for 95% confidence intervals; and (iv) columns B/SE and MSE report, respectively, simulation bias relative to simulation standard error and simulation mean squared error of $\hat{\theta}_n(h_n)$.

the ranking across inference procedures in terms of coverage is in perfect agreement with our theoretical predictions.

8. CONCLUSION

This paper has developed "small bandwidth" asymptotic results for a large class of two-step kernel-based semiparametric estimators. Our first main result, Theorem 1, differs from those obtained in earlier work on semiparametric two-step estimators by accommo-

TABLE VI

SIMULATION RESULTS FOR MODELS M21–M25 ($n = 1{,}000$, $B = 1{,}000$, $S = 1{,}000$)[a]

	BW		CR			IL				
	h_n	c	E	N	P	E	N	P	B/SE	MSE
M21: $d = 1, P = 2$	0.018	0.5	0.954	0.946	0.935	0.066	0.063	0.066	−0.146	1.008
	0.022	0.6	0.956	0.949	0.932	0.066	0.063	0.066	−0.106	1.010
	0.025	0.7	0.957	0.948	0.932	0.066	0.063	0.066	−0.072	1.007
	0.029	0.8	0.957	0.948	0.934	0.066	0.063	0.066	−0.042	1.004
	0.033	0.9	0.958	0.947	0.938	0.065	0.063	0.065	−0.014	1.002
$h_{\mathtt{opt}} =$	0.036	1.0	0.958	0.950	0.940	0.065	0.063	0.065	0.014	1.000
	0.040	1.1	0.958	0.951	0.940	0.065	0.063	0.065	0.043	1.001
M22: $d = 1, P = 4$	0.027	0.5	0.961	0.950	0.936	0.066	0.063	0.066	−0.126	0.995
	0.033	0.6	0.961	0.949	0.932	0.066	0.064	0.066	−0.101	1.003
	0.038	0.7	0.959	0.947	0.932	0.066	0.064	0.066	−0.087	1.004
	0.044	0.8	0.958	0.948	0.931	0.066	0.064	0.066	−0.078	1.004
	0.049	0.9	0.957	0.948	0.936	0.066	0.064	0.066	−0.074	1.002
$h_{\mathtt{opt}} =$	0.055	1.0	0.956	0.949	0.938	0.066	0.064	0.066	−0.071	1.000
	0.060	1.1	0.956	0.947	0.939	0.065	0.064	0.065	−0.070	0.995
M23: $d = 2, P = 2$	0.033	0.5	0.013	0.473	0.874	0.053	0.053	0.053	−2.055	2.946
	0.040	0.6	0.192	0.676	0.894	0.055	0.054	0.055	−1.484	1.861
	0.046	0.7	0.438	0.775	0.904	0.056	0.054	0.056	−1.174	1.395
	0.053	0.8	0.606	0.821	0.912	0.056	0.055	0.056	−0.992	1.175
	0.060	0.9	0.709	0.851	0.910	0.056	0.055	0.056	−0.886	1.060
$h_{\mathtt{opt}} =$	0.066	1.0	0.766	0.876	0.909	0.056	0.055	0.056	−0.818	1.000
	0.073	1.1	0.802	0.884	0.908	0.056	0.055	0.056	−0.777	0.962
M24: $d = 2, P = 4$	0.043	0.5	0.001	0.341	0.848	0.052	0.053	0.052	−2.395	3.862
	0.052	0.6	0.078	0.619	0.882	0.055	0.054	0.055	−1.643	2.233
	0.060	0.7	0.337	0.775	0.898	0.056	0.055	0.056	−1.229	1.545
	0.069	0.8	0.557	0.838	0.909	0.057	0.055	0.057	−0.983	1.219
	0.078	0.9	0.687	0.865	0.918	0.057	0.055	0.057	−0.854	1.067
$h_{\mathtt{opt}} =$	0.086	1.0	0.757	0.875	0.917	0.057	0.055	0.057	−0.788	1.000
	0.095	1.1	0.792	0.883	0.912	0.057	0.055	0.057	−0.767	0.975
M25: $d = 3, P = 4$	0.060	0.5	0.000	0.000	0.000	0.022	0.033	0.022	−18.517	23.105
	0.072	0.6	0.000	0.000	0.122	0.033	0.042	0.033	−8.922	8.607
	0.084	0.7	0.000	0.001	0.665	0.041	0.046	0.041	−5.296	3.754
	0.096	0.8	0.000	0.050	0.800	0.047	0.049	0.047	−3.587	1.982
	0.108	0.9	0.000	0.228	0.825	0.049	0.050	0.049	−2.723	1.284
$h_{\mathtt{opt}} =$	0.120	1.0	0.020	0.351	0.812	0.051	0.051	0.051	−2.316	1.000
	0.132	1.1	0.082	0.406	0.775	0.052	0.052	0.052	−2.160	0.906

[a](i) Columns under BW report grid of bandwidths and multiplicative factor c relative to $h_{\mathtt{opt}}$; (ii) columns under CR report coverage error for 95% confidence intervals; (iii) columns under IL report average interval length for 95% confidence intervals; and (iv) columns B/SE and MSE report, respectively, simulation bias relative to simulation standard error and simulation mean squared error of $\hat{\theta}_n(h_n)$.

dating a nonnegligible bias. A noteworthy feature of the assumptions of this theorem is that reliance on a commonly employed stochastic equicontinuity condition is avoided. The second main result, Theorem 2, shows that the bootstrap provides an automatic method of correcting for the bias even when it is nonnegligible.

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 1 and 2 along the lines of Romano and Shaikh (2012) and Einmahl and Mason (2005), respectively.

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., sieve estimators of M-regression functions, possibly after model selection as in Belloni, Chernozhukov, Fernández-Val, and Hansen (2017) and references therein). In this paper, we focus on kernel-based estimators because of their analytical tractability, but we conjecture that our results can be extended to cover other nonparametric first-step estimators. In future work, we intend to attempt to substantiate this conjecture.

APPENDIX A: PROOFS

A.1. *Proof of Theorem 1*

The proof is elementary:

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_0 - \mathcal{J}_n\mathcal{B}_n) &= \big[\mathcal{J}_0 + o(1)\big]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_n\big(z_i, \hat{\gamma}_n^{(i)}\big) - \mathcal{B}_n\big] + o_{\mathbb{P}}(1)\\ &= \big[\mathcal{J}_0 + o(1)\big]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_n(z_i, \gamma_n) + \bar{G}_n\big(\hat{\gamma}_n^{(i)}\big) - \bar{G}_n(\gamma_n) - \mathcal{B}_n\big] + o_{\mathbb{P}}(1)\\ &\rightsquigarrow \mathcal{N}\big(0, \mathcal{J}_0\Omega_0\mathcal{J}_0'\big),\end{aligned}$$

where the first equality uses Condition AL and (4), the second equality uses Condition AS, and the last line uses Condition AN.

A.2. *Proof of Theorem 2*

The proof is elementary:

$$\begin{aligned}&\sqrt{n}\big(\hat{\theta}_n^* - \hat{\theta}_n - \mathcal{J}_n^*\mathcal{B}_n^*\big)\\ &\quad= \big[\mathcal{J}_0^* + o(1)\big]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_n^*\big(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}\big) - \mathcal{B}_n^*\big] + o_{\mathbb{P}}(1)\\ &\quad= \big[\mathcal{J}_0^* + o(1)\big]\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_n^*\big(z_{i,n}^*, \hat{\gamma}_n\big) + \bar{G}_n^*\big(\hat{\gamma}_n^{*,(i)}\big) - \bar{G}_n^*(\hat{\gamma}_n) - \mathcal{B}_n^*\big] + o_{\mathbb{P}}(1)\\ &\quad\rightsquigarrow_{\mathbb{P}} \mathcal{N}\big(0, \mathcal{J}_0^*\Omega_0^*\mathcal{J}_0^{*\prime}\big),\end{aligned}$$

where the first equality uses Condition AL* and (9), the second equality uses Condition AS*, and the last line uses Condition AN*.

A.3. *Proof of Lemma 1*

Using (iv), (vi), and $\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n) \to_{\mathbb{P}} \dot{G}_0'W_0\dot{G}_0 > 0$, we have

$$(\hat{\mathcal{J}}_n - \mathcal{J}_n)\hat{G}_n(\theta_0, \hat{\gamma}_n) = o_{\mathbb{P}}\big(n^{-1/2}\big), \qquad \hat{\mathcal{J}}_n = -\big[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\big]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n.$$

As a consequence, it suffices to show that $\hat{\theta}_n - \theta_0 - \hat{\mathcal{J}}_n \hat{G}_n(\theta_0, \hat{\gamma}_n) = o_{\mathbb{P}}(n^{-1/2})$. To do so, let

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

Because $\dot{G}(\hat{\gamma}_n)' \hat{W}_n \dot{G}(\hat{\gamma}_n) \to_{\mathbb{P}} \dot{G}_0' W_0 \dot{G}_0 > 0$ and

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

it suffices to show that $L_n(\hat{\theta}_n) = o_{\mathbb{P}}(n^{-1/2})$.

If $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$, then

$$\big\| G(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\hat{\theta}_n - \theta_0) \big\| = \|\hat{\theta}_n - \theta_0\|^{\rho/2} o_{\mathbb{P}}(1) = o_{\mathbb{P}}\big(n^{-1/2}\big)$$

and

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

by (ii) and (vii), respectively. As a consequence, by the triangle inequality,

$$\begin{aligned} \big\| L_n(\hat{\theta}_n) \big\| &\le \big\| \dot{G}(\hat{\gamma}_n)' \hat{W}_n \big\| \big\| \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n) \big\| \\ &\quad + \big\| \dot{G}(\hat{\gamma}_n)' \hat{W}_n \big\| \big\| G(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\hat{\theta}_n - \theta_0) \big\| \\ &\quad + \big\| \dot{G}(\hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \big\| \\ &= o_{\mathbb{P}}\big(n^{-1/2}\big), \end{aligned}$$

where the equality uses $\|\dot{G}(\hat{\gamma}_n)' \hat{W}_n\| = O_{\mathbb{P}}(1)$ and $\dot{G}(\hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}(n^{-1/2})$. The proof can therefore be completed by showing that $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$.

Proof of $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$. Because $\hat{\theta}_n - \theta_0 = o_{\mathbb{P}}(1)$, $\hat{W}_n^{1/2} \dot{G}(\hat{\gamma}_n) - W_0^{1/2} \dot{G}_0 = o_{\mathbb{P}}(1)$, and $\dot{G}_0' W_0 \dot{G}_0 > 0$, condition (ii) implies that

$$\|\hat{\theta}_n - \theta_0\| \le \big\| \hat{W}_n^{1/2} \big[G(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n)\big] \big\| O_{\mathbb{P}}(1),$$

so it suffices to show that $\hat{W}_n^{1/2}[G(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n)] \le O_{\mathbb{P}}(n^{-1/\rho}) + \|\hat{\theta}_n - \theta_0\| o_{\mathbb{P}}(1)$.

Using (i) and (iv), we have $\hat{W}_n^{1/2} \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/\rho})$ because

$$\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \le \hat{G}_n(\theta_0, \hat{\gamma}_n)' \hat{W}_n \hat{G}_n(\theta_0, \hat{\gamma}_n) + o_{\mathbb{P}}\big(n^{-1}\big) = O_{\mathbb{P}}\big(n^{-2/\rho}\big).$$

Also, using $\hat{\theta}_n - \theta_0 = o_{\mathbb{P}}(1)$ and (iii),

$$\begin{aligned} &\big\| \hat{W}_n^{1/2} \big[\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\big] \big\| \\ &\quad = o_{\mathbb{P}}\big(n^{-1/\rho}\big) + \|\hat{\theta}_n - \theta_0\| o_{\mathbb{P}}(1), \end{aligned}$$

so

$$\begin{aligned} &\big\| \hat{W}_n^{1/2} \big[G(\hat{\theta}_n, \hat{\gamma}_n) - G(\theta_0, \hat{\gamma}_n)\big] \big\| \\ &\quad \le \big\| \hat{W}_n^{1/2} \big[\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\big] \big\| \end{aligned}$$

$$
\begin{aligned}
&+ \bigl\|\hat{W}_n^{1/2}\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)\bigr\| + \bigl\|\hat{W}_n^{1/2}\hat{G}_n(\theta_0, \hat{\gamma}_n)\bigr\| \\
&= O_{\mathbb{P}}\bigl(n^{-1/\rho}\bigr) + \|\hat{\theta}_n - \theta_0\| o_{\mathbb{P}}(1),
\end{aligned}
$$

where the inequality uses the triangle inequality and the equality uses (iv).

A.4. *Verifying* $\dot{G}(\hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}(n^{-1/2})$

Suppose the conditions of Lemma 1 are satisfied, with the possible exception of

$$
\dot{G}(\hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}\bigl(n^{-1/2}\bigr). \tag{A-1}
$$

Because $\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n) \to_{\mathbb{P}} \dot{G}_0'W_0\dot{G}_0 > 0$, (A-1) holds provided

$$
\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = o_{\mathbb{P}}\bigl(n^{-1}\bigr).
$$

To give conditions under which the latter holds, let $\tilde{\theta}_n = \hat{\theta}_n + \hat{\mathcal{J}}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)$, which satisfies $\tilde{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$ because $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$ and $\hat{W}_n^{1/2}\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) = O_{\mathbb{P}}(n^{-1/\rho})$.

Defining

$$
R_n = \hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - \dot{G}(\hat{\gamma}_n)(\tilde{\theta}_n - \hat{\theta}_n),
$$

and using the fact that θ_0 is an interior point of Θ, we have

$$
\begin{aligned}
&\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\quad\le \hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n) + o_{\mathbb{P}}\bigl(n^{-1}\bigr) \\
&\quad= \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\qquad- \hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\qquad+ 2R_n'\bigl[\hat{W}_n - \hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\bigr]\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\qquad+ R_n'\hat{W}_nR_n + o_{\mathbb{P}}\bigl(n^{-1}\bigr),
\end{aligned}
$$

which rearranges as

$$
\begin{aligned}
&\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) \\
&\quad\le 2R_n'\bigl[\hat{W}_n - \hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\bigr]\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) + R_n'\hat{W}_nR_n + o_{\mathbb{P}}\bigl(n^{-1}\bigr) \\
&\quad= 2R_n'\bigl[\hat{W}_n - \hat{W}_n\dot{G}(\hat{\gamma}_n)\bigl[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\bigr]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n\bigr]\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) + o_{\mathbb{P}}\bigl(n^{-1}\bigr),
\end{aligned}
$$

where the equality uses

$$
\begin{aligned}
\|R_n\| &\le \bigl\|\hat{G}_n(\hat{\theta}_n, \hat{\gamma}_n) - G(\hat{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)\bigr\| \\
&\quad+ \bigl\|\hat{G}_n(\tilde{\theta}_n, \hat{\gamma}_n) - G(\tilde{\theta}_n, \hat{\gamma}_n) - \hat{G}_n(\theta_0, \hat{\gamma}_n) + G(\theta_0, \hat{\gamma}_n)]\bigr\| \\
&\quad+ \|\hat{\theta}_n - \theta_0\|^2 O_{\mathbb{P}}(1) + \|\tilde{\theta}_n - \theta_0\|^2 O_{\mathbb{P}}(1) \\
&= o_{\mathbb{P}}\bigl(n^{-1/2}\bigr).
\end{aligned}
$$

The desired result therefore follows if either

$$\hat{W}_n - \hat{W}_n\dot{G}(\hat{\gamma}_n)\big[\dot{G}(\hat{\gamma}_n)'\hat{W}_n\dot{G}(\hat{\gamma}_n)\big]^{-1}\dot{G}(\hat{\gamma}_n)'\hat{W}_n = O_{\mathbb{P}}\big(n^{-1/2}\big)$$

or $R_n = o_{\mathbb{P}}(n^{1/\rho-1})$. The latter condition is satisfied if $\rho < 3$ and if, for every $\delta_n = O(n^{-1/\rho})$,

$$\sup_{\|\theta-\theta_0\|\le\delta_n} \big\|\hat{G}_n(\theta,\hat{\gamma}_n) - G(\theta,\hat{\gamma}_n) - \hat{G}_n(\theta_0,\hat{\gamma}_n) + G(\theta_0,\hat{\gamma}_n)\big\| = o_{\mathbb{P}}\big(n^{1/\rho-1}\big).$$

The former condition is satisfied if either g is of dimension d_θ or if

$$\hat{W}_n = \dot{G}(\hat{\gamma}_n)\dot{G}(\hat{\gamma}_n)' + O_{\mathbb{P}}\big(n^{-1/2}\big).$$

A.5. *Proof of Lemma 2*

By construction, $n^{-1/2}\sum_{i=1}^n[\bar{g}_n(z_i,\hat{\gamma}_n^{(i)}) - \bar{G}_n(\hat{\gamma}_n^{(i)}) - \bar{g}_n(z_i,\gamma_n) + \bar{G}_n(\gamma_n)]$ has mean zero, so it suffices to show that its variance converges to zero. Using the decomposition

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

and Hoeffding's theorem for U-statistics, we have

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\bar{g}_n\big(z_i,\hat{\gamma}_n^{(i)}\big) - \bar{G}_n\big(\hat{\gamma}_n^{(i)}\big) - \bar{g}_n(z_i,\gamma_n) + \bar{G}_n(\gamma_n)\big]\right)\\
&\quad= \frac{1}{n}O\big(\mathbb{V}\big(g_{n,\gamma}(z_i)\big[\hat{\gamma}_n^j-\gamma_n\big]\big)\big) + \frac{1}{n^2}O\big(\mathbb{V}\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j-\gamma_n,\hat{\gamma}_n^k-\gamma_n\big]\big)\big)\\
&\qquad+ \frac{1}{n^2}O\big(\mathbb{V}\big[E\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j-\gamma_n,\hat{\gamma}_n^j-\gamma_n\big]|z_i\big)\big]\big) + \frac{1}{n^3}O\big(\mathbb{V}\big(g_{n,\gamma\gamma}(z_i)\big[\hat{\gamma}_n^j-\gamma_n,\hat{\gamma}_n^j-\gamma_n\big]\big)\big)\\
&\quad= o(1),
\end{aligned}$$

where the last equality uses the assumptions displayed in the statement of the lemma.

A.6. *Proof of Lemma 3*

Because

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

it follows from Hoeffding's theorem for U-statistics that if the assumptions displayed in the statement of the lemma are satisfied, then

$$\mathbb{V}(\sqrt{n}\hat{\mathcal{B}}_n)=\frac{1}{n^2}O\big(\mathbb{V}\big(G_{n,\gamma\gamma}\big[\hat{\gamma}_n^{i}-\gamma_n,\hat{\gamma}_n^{i}-\gamma_n\big]\big)\big)+\frac{1}{n}O\big(\mathbb{V}\big(G_{n,\gamma\gamma}\big[\hat{\gamma}_n^{i}-\gamma_n,\hat{\gamma}_n^{j}-\gamma_n\big]\big)\big)=o(1),$$

implying in particular that $\sqrt{n}(\hat{\mathcal{B}}_n-\mathbb{E}\hat{\mathcal{B}}_n)=o_{\mathbb{P}}(1)$.

If also (10) is satisfied, then Condition AN holds with $\Omega_0=\mathbb{V}[\psi_0(z)]$ and any $\mathcal{B}_n=\mathbb{E}\hat{\mathcal{B}}_n+o(n^{-1/2})$ because

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g_n(z_i,\gamma_n)+\bar{G}_n\big(\hat{\gamma}_n^{(i)}\big)-\bar{G}_n(\gamma_n)-\mathcal{B}_n\big]\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n(z_i)+\sqrt{n}(\hat{\mathcal{B}}_n-\mathbb{E}\hat{\mathcal{B}}_n)+\sqrt{n}(\mathbb{E}\hat{\mathcal{B}}_n-\mathcal{B}_n)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_0(z_i)+o_{\mathbb{P}}(1)\rightsquigarrow\mathcal{N}(0,\Omega_0).
\end{aligned}$$

A.7. *Proof of Lemma 4*

Using (iv*), (vi*), and $\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\dot{G}(\hat{\gamma}_n^*)\to_{\mathbb{P}}\dot{G}_0'W_0\dot{G}_0>0$, we have

$$\big(\hat{\mathcal{J}}_n^*-\mathcal{J}_n\big)\hat{G}_n^*\big(\hat{\theta}_n,\hat{\gamma}_n^*\big)=o_{\mathbb{P}}\big(n^{-1/2}\big),\qquad \hat{\mathcal{J}}_n^*=-\big[\dot{G}\big(\hat{\gamma}_n^*\big)'\hat{W}_n^*\dot{G}\big(\hat{\gamma}_n^*\big)\big]^{-1}\dot{G}\big(\hat{\gamma}_n^*\big)'\hat{W}_n^*.$$

As a consequence, it suffices to show that $\hat{\theta}_n^*-\hat{\theta}_n-\hat{\mathcal{J}}_n^*\hat{G}_n^*(\hat{\theta}_n,\hat{\gamma}_n^*)=o_{\mathbb{P}}(n^{-1/2})$. To do so, let

$$L_n^*(\theta)=\dot{G}\big(\hat{\gamma}_n^*\big)'\hat{W}_n^*\big[\hat{G}_n^*\big(\hat{\theta}_n,\hat{\gamma}_n^*\big)+\dot{G}\big(\hat{\gamma}_n^*\big)(\theta-\hat{\theta}_n)\big].$$

Because $\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\dot{G}(\hat{\gamma}_n^*) \to_{\mathbb{P}} \dot{G}_0'W_0\dot{G}_0 > 0$ and

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

it suffices to show that $L_n^*(\hat{\theta}_n^*) = o_{\mathbb{P}}(n^{-1/2})$.

Because $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$,

$$\begin{aligned}\big\|G(\hat{\theta}_n, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*) - \dot{G}(\hat{\gamma}_n)(\hat{\theta}_n - \theta_0)\big\| &= \|\hat{\theta}_n - \theta_0\|^{\rho/2}o_{\mathbb{P}}(1)\\ &= o_{\mathbb{P}}\big(n^{-1/2}\big)\end{aligned}$$

and

$$\big\|\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) + G(\theta_0, \hat{\gamma}_n^*)\big\| = o_{\mathbb{P}}\big(n^{-1/2}\big)$$

by (ii*) and (vii*), respectively. If also $\hat{\theta}_n^* - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$, then

$$\big\|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*) - \dot{G}(\hat{\gamma}_n^*)(\hat{\theta}_n^* - \theta_0)\big\| = \big\|\hat{\theta}_n^* - \theta_0\big\|^{\rho/2}o_{\mathbb{P}}(1) = o_{\mathbb{P}}\big(n^{-1/2}\big)$$

and

$$\big\|\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) + G(\theta_0, \hat{\gamma}_n^*)\big\| = o_{\mathbb{P}}\big(n^{-1/2}\big)$$

by (ii*) and (vii*), respectively. As a consequence, by the triangle inequality,

$$\begin{aligned}\big\|L_n^*(\hat{\theta}_n^*)\big\| \le{}& \big\|\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\big\|\big\|\hat{G}_n^*(\hat{\theta}_n, \hat{\gamma}_n^*) - G(\hat{\theta}_n, \hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) + G(\theta_0, \hat{\gamma}_n^*)\big\|\\ &+ \big\|\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\big\|\big\|G(\hat{\theta}_n, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*) - \dot{G}(\hat{\gamma}_n)(\hat{\theta}_n - \theta_0)\big\|\\ &+ \big\|\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\big\|\big\|\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - \hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) + G(\theta_0, \hat{\gamma}_n^*)\big\|\\ &+ \big\|\dot{G}(\hat{\gamma}_n)'\hat{W}_n\big\|\big\|G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*) - \dot{G}(\hat{\gamma}_n^*)(\hat{\theta}_n^* - \theta_0)\big\|\\ &+ \big\|\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)\big\|\\ ={}& o_{\mathbb{P}}\big(n^{-1/2}\big),\end{aligned}$$

where the equality uses $\|\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\| = O_{\mathbb{P}}(1)$ and $\dot{G}(\hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = o_{\mathbb{P}}(n^{-1/2})$. The proof can therefore be completed by showing that $\hat{\theta}_n^* - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$.

Proof of $\hat{\theta}_n^* - \theta_0 = O_{\mathbb{P}}(n^{-1/\rho})$. Because $\hat{\theta}_n^* - \theta_0 = o_{\mathbb{P}}(1)$ and $\hat{W}_n^{1/2}\dot{G}(\hat{\gamma}_n^*) - W_0^{1/2}\dot{G}_0 = o_{\mathbb{P}}(1)$, condition (ii*) implies that

$$\big\|\hat{\theta}_n^* - \theta_0\big\| \le \big\|\hat{W}_n^{*1/2}\big[G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*)\big]\big\|O_{\mathbb{P}}(1),$$

so it suffices to show that $\hat{W}_n^{*1/2}[G(\hat{\theta}_n^*, \hat{\gamma}_n^*) - G(\theta_0, \hat{\gamma}_n^*)] \le O_{\mathbb{P}}(n^{-1/\rho}) + \|\hat{\theta}_n^* - \theta_0\|o_{\mathbb{P}}(1)$.

Using (i*) and (iv*), we have $\hat{W}_n^{*1/2}\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) = O_{\mathbb{P}}(n^{-1/\rho})$ because

$$\begin{aligned}\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\hat{\theta}_n^*, \hat{\gamma}_n^*) &\le \hat{G}_n^*(\theta_0, \hat{\gamma}_n^*)'\hat{W}_n^*\hat{G}_n^*(\theta_0, \hat{\gamma}_n^*) + o_{\mathbb{P}}\big(n^{-1}\big)\\ &= O_{\mathbb{P}}\big(n^{-2/\rho}\big).\end{aligned}$$

Also, using $\hat{\theta}_n^* - \theta_0 = o_{\mathbb{P}}(1)$ and (iii*),

$$\begin{aligned}&\big\|\hat{W}_n^{*1/2}\big[\hat{G}_n^*\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big) - G\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big) - \hat{G}_n^*\big(\theta_0, \hat{\gamma}_n^*\big) + G\big(\theta_0, \hat{\gamma}_n^*\big)\big]\big\| \\ &\quad = o_{\mathbb{P}}\big(n^{-1/\rho}\big) + \big\|\hat{\theta}_n^* - \theta_0\big\| o_{\mathbb{P}}(1),\end{aligned}$$

so

$$\begin{aligned}&\big\|\hat{W}_n^{*1/2}\big[G\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big) - G\big(\theta_0, \hat{\gamma}_n^*\big)\big]\big\| \\ &\quad \le \big\|\hat{W}_n^{*1/2}\big[\hat{G}_n^*\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big) - G\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big) - \hat{G}_n^*\big(\theta_0, \hat{\gamma}_n^*\big) + G\big(\theta_0, \hat{\gamma}_n^*\big)\big]\big\| \\ &\qquad + \big\|\hat{W}_n^{*1/2}\hat{G}_n^*\big(\hat{\theta}_n^*, \hat{\gamma}_n^*\big)\big\| + \big\|\hat{W}_n^{*1/2}\hat{G}_n^*\big(\theta_0, \hat{\gamma}_n^*\big)\big\| \\ &\quad = O_{\mathbb{P}}\big(n^{-1/\rho}\big) + \|\hat{\theta}_n - \theta_0\| o_{\mathbb{P}}(1),\end{aligned}$$

where the inequality uses the triangle inequality and the equality uses (iv*).

A.8. *Proof of Lemma 5*

By construction,

$$\mathbb{E}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\bar{g}_n^*\big(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}\big) - \bar{G}_n^*\big(\hat{\gamma}_n^{*,(i)}\big) - \bar{g}_n^*\big(z_{i,n}^*, \hat{\gamma}_n\big) + \bar{G}_n^*(\hat{\gamma}_n)\big]\right) = 0.$$

Moreover, using the decomposition

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\bar{g}_n^*\big(z_{i,n}^*, \hat{\gamma}_n^{*,(i)}\big) - \bar{G}_n^*\big(\hat{\gamma}_n^{*,(i)}\big) - \bar{g}_n^*\big(z_{i,n}^*, \hat{\gamma}_n\big) + \bar{G}_n^*(\hat{\gamma}_n)\big] \\ &\quad = \frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^n\sum_{j=1, j\neq i}^n\big(g_{n,\gamma}^*\big(z_{i,n}^*\big)\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n\big] - G_{n,\gamma}^*\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n\big]\big) \\ &\qquad + \frac{1}{2\sqrt{n}(n-1)^2}\sum_{i=1}^n\sum_{j=1, j\neq i}^n\big(g_{n,\gamma\gamma}^*\big(z_{i,n}^*\big)\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n\big] \\ &\qquad - G_{n,\gamma\gamma}^*\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,j} - \hat{\gamma}_n\big]\big) \\ &\qquad + \frac{1}{2\sqrt{n}(n-1)^2}\sum_{i=1}^n\sum_{j=1, j\neq i}^n\sum_{k=1, k\notin\{i,j\}}^n\big(g_{n,\gamma\gamma}^*\big(z_{i,n}^*\big)\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,k} - \hat{\gamma}_n\big] \\ &\qquad - G_{n,\gamma\gamma}^*\big[\hat{\gamma}_n^{*,j} - \hat{\gamma}_n, \hat{\gamma}_n^{*,k} - \hat{\gamma}_n\big]\big),\end{aligned}$$

and proceeding as in the proof of Lemma 2, it follows from the assumptions displayed in the statement of Lemma 5 that

$$\mathbb{V}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\bar{g}_n\big(z_i, \hat{\gamma}_n^{(i)}\big) - \bar{G}_n\big(\hat{\gamma}_n^{(i)}\big) - \bar{g}_n(z_i, \gamma_n) + \bar{G}_n(\gamma_n)\big]\right) = o_{\mathbb{P}}(1).$$

A.9. *Proof of Lemma 6*

Because

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} G^*_{n,\gamma\gamma}\big[\hat\gamma_n^{*,(i)}-\hat\gamma_n,\hat\gamma_n^{*,(i)}-\hat\gamma_n\big]=\frac{1}{n-1}\sum_{i=1}^{n} G^*_{n,\gamma\gamma}\big[\hat\gamma_n^{*,i}-\hat\gamma_n,\hat\gamma_n^{*,i}-\hat\gamma_n\big] + \frac{n-2}{(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} G^*_{n,\gamma\gamma}\big[\hat\gamma_n^{*,i}-\hat\gamma_n,\hat\gamma_n^{*,j}-\hat\gamma_n\big],$$

it follows from Hoeffding's theorem for U-statistics that if the assumptions displayed in the statement of the lemma are satisfied, then

$$\sqrt{n}\big(\hat{\mathcal{B}}^*_n-\mathbb{E}\hat{\mathcal{B}}^*_n\big)=\sqrt{n}\big(\hat{\mathcal{B}}^*_n-\mathbb{E}^*\hat{\mathcal{B}}^*_n\big)+\sqrt{n}\big(\mathbb{E}^*\hat{\mathcal{B}}^*_n-\mathbb{E}\hat{\mathcal{B}}^*_n\big)=o_{\mathbb{P}}(1),$$

and therefore

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g^*_n\big(z^*_{i,n},\hat\gamma_n\big)+\bar G^*_n\big(\hat\gamma_n^{(i)}\big)-\bar G^*_n(\hat\gamma_n)-\mathcal{B}^*_n\big]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^*_n\big(z^*_{i,n}\big)+o_{\mathbb{P}}(1)$$

for any $\mathcal{B}^*_n=\mathbb{E}\hat{\mathcal{B}}^*_n+o(n^{-1/2})$. If also (12) is satisfied, then

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^*_n\big(z^*_{i,n}\big)&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n\big(z^*_{i,n}\big)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_n(z_i)+o_{\mathbb{P}}(1)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_0\big(z^*_{i,n}\big)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_0(z_{i,n})+o_{\mathbb{P}}(1)\rightsquigarrow_{\mathbb{P}}\mathcal{N}(0,\Omega_0),\end{aligned}$$

where the second equality uses (10).

REFERENCES

ABADIE, A., AND G. W. IMBENS (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. [956]

——— (2008): "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 76, 1537–1557. [956]

ANDREWS, D. W. K. (1994a): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72. [963]

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. New York: North Holland, 2247–2294. [957,958,963,974]

BELLONI, A., V. CHERNOZHUKOV, I. FERNÁNDEZ-VAL, AND C. HANSEN (2017): "Program Evaluation With High-Dimensional Data," *Econometrica*, 85, 233–298. [987]

BICKEL, P. J., AND B. LI (2006): "Regularization in Statistics," *Test*, 15, 271–344. [957,969]

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. [972]

CANAY, I. A., J. P. ROMANO, AND A. M. SHAIKH (2017): "Randomization Tests Under an Approximate Symmetry Condition," *Econometrica*, 85, 1013–1030. [969,970]

CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2010): "Robust Data-Driven Inference for Density-Weighted Average Derivatives," *Journal of the American Statistical Association*, 105, 1070–1083. [956]

——— (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives (With Discussion and Rejoinder)," *Journal of the American Statistical Association*, 108, 1243–1268. [956,965]

——— (2014a): "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," *Econometric Theory*, 30, 176–200. [956]

——— (2014b): "Bootstrapping Density-Weighted Average Derivatives," *Econometric Theory*, 30, 1135–1164. [956]

CATTANEO, M. D., AND M. JANSSON (2018): "Supplement to 'Kernel-Based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency'," *Econometrica Supplemental Material*, 86, https://doi.org/10.3982/ECTA12701. [958]

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, Vol. 6B, ed. by J. J. Heckman and E. E. Leamer. New York: North Holland, 5549–5632. [957,958]

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): "Estimation of Semiparametric Models When the Criterion Function Is not Smooth," *Econometrica*, 71, 1591–1608. [957,974,980]

CHENG, G., AND J. C. HUANG (2010): "Bootstrap Consistency for General Semiparametric M-Estimation," *Annals of Statistics*, 38, 2884–2915. [957]

CHERNOZHUKOV, V., J. C. ESCANCIANO, H. ICHIMURA, AND W. K. NEWEY (2016): "Locally Robust Semiparametric Estimation". arXiv:1608.00033. [957,965]

EINMAHL, U., AND D. M. MASON (2005): "Uniform in Bandwidth Consistency of Kernel-Type Function Estimators," *Annals of Statistics*, 33, 1380–1403. [987]

FAN, J., AND I. GIJBELS (1997): *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall. [959]

HAHN, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12, 187–197. [972]

HALL, P., AND J. L. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891–916. [972]

HAUSMAN, J. A., AND W. K. NEWEY (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63, 1445–1476. [964]

IBRAGIMOV, R., AND U. K. MÜLLER (2010): "t-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468. [969,970]

ICHIMURA, H., AND P. E. TODD (2007): "Implementing Nonparametric and Semiparametric Estimators," in *Handbook of Econometrics*, Vol. 6B, ed. by J. J. Heckman and E. E. Leamer. New York: North Holland, 5369–5468. [957,958]

MAMMEN, E. (1989): "Asymptotics With Increasing Dimension for Robust Regression With Applications to the Bootstrap," *Annals of Statistics*, 17, 382–400. [956]

NEWEY, W. K. (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. [958,963,975]

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253. [977]

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. New York: North Holland, 2111–2245. [957-959]

NEWEY, W. K., F. HSIEH, AND J. M. ROBINS (2004): "Twicing Kernels and a Small Bias Property of Semiparametric Estimators," *Econometrica*, 72, 947–962. [965]

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [972,973]

POLITIS, D. N., AND J. P. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050. [969]

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430. [956,964]

RADULOVIC, D. (1998): "Can We Bootstrap Even if CLT Fails?" *Journal of Theoretical Probability*, 11, 813–830. [970]

ROBINS, J., L. LI, R. MUKHERJEE, E. TCHETGEN, AND A. VAN DER VAART (2017): "Minimax Estimation of a Functional on a Structured High-Dimensional Model," *Annals of Statistics*, 45, 1951–1987. [957]

ROBINS, J., L. LI, E. TCHETGEN, AND A. VAN DER VAART (2008): "Higher Order Influence Functions and Minimax Estimation of Nonlinear Functionals," in *Probability and Statistics: Essays in Honor of David A. Freedman*, ed. by D. Nolan and T. Speed. Beachwood, OH: Institute of Mathematical Statistics, 335–421. [957]

——— (2016): "Asymptotic Normality of Quadratic Estimators," *Stochastic Processes and their Applications*, 126, 3733–3759. [957]

ROMANO, J. P., AND A. M. SHAIKH (2012): "On the Uniform Asymptotic Validity of Subsampling and the Bootstrap," *Annals of Statistics*, 40, 2798–2822. [987]

Co-editor Elie Tamer handled this manuscript.

Manuscript received 30 July, 2014; final version accepted 5 December, 2017; available online 17 January, 2018.